DOCUMENT RESUME

ED 223 711 TM 820 858

AUTHOR	De Avila, Edward; And Others
TITLE	A Neo-Piagetian Approach to Test Bias: Final Report.
INSTITUTION	De Avila, Duncan and Associates, Inc., Larkspur, CA.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	31 Mar 82
GRANT	NIE-G-79-0155
NOTE	203p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC09 Plus Postage.
DESCRIPTORS	Cognitive Processes; Cognitive Style; *Cognitive Tests; *Developmental Stages; Elementary Education; Ethnic Groups; Intentional Learning; Performance Factors; *Test Bias; *Test Coaching; *Test Wiseness
IDENTIFIERS	*Neo Piagetian Theory; *Raven Progressive Matrices

ABSTRACT

This project examined the hypothesis that different background experiences associated with cultural grouping may lead to differences in test-taking strategies which result in score differences extraneous to the abilities the test is intended to measure. Its purposes were to confirm (or disconfirm) the cultural differences hypothesis and to provide a systematic basis for reducing this potential source of test bias and invalidity. Subjects were 810 Anglo, Black, and Mexican-American students in grades 2, 4, and 6. The test used was the Raven Progressive Matrices. The relative level of culturally related bias was predicted for each item a priori, based on level of complexity. Results indicated that although group differences on the test are related to developmental level, they are also related to test-taking skills. Test-taking skills are a major source of variation; they are learned and can be strengthened through exposure to specific requirements for the test. These results indicate that an important source of bias is as much in the overall testing procedure as in the test itself, and challenge the assumption that all children approach and solve a test-taking task in the same way. (Author/PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED223711

A Neo-Piagetian Approach to Test Bias:

Final Report

to

National Institute of Education
March 31, 1982

De Avila, Duncan & Associates, Inc.
San Rafael, California

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

TM 820 858

FB

DE AVILA, DUNCAN & ASSOCIATES, INC.
P.O. Box 770
Larkspur, Ca 94939
Phone: 415 459-0212

A Neo-Piagetian Approach to Test Bias:
Final Report
to
National Institute of Education

March 31, 1982

De Avila, Duncan & Associates, Inc., San Rafael, California

"This product has been produced with support from Grant number NIE-G-79-0155 which was funded by the Department of Health, Education, and Welfare, National Institute of Education. The contents of this publication do not necessarily reflect the views or policies of the National Institute of Education, or the Department of Health, Education, and Welfare, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government."

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES vi

I. INTRODUCTION AND THEORETICAL POSITION 1

Information Processing Capacity and Task Complexity . . 4
Task Complexity and Culture-Loading 7
Information Processing Capacity and Cognitive Style . . 9
Culture-Loading and Test Bias 10

II. DESIGN OF THE STUDY 12

Methodology 12
Subjects 13
Procedure 15
Criterion Tests 16
  Information Processing Capacity 16
  Cognitive Style 17
  Sensitivity to Misleading Cues. 17
  Developmental Level 18
Analytic Intelligence 19
Training 19

III. RESULTS: EFFECTS OF TRAINING 27

Test Performance and Effects of Training 27

Relationship Among Tests 28
Intercorrelations Among Tests 31
Factor Analysis 34
Information Processing Capacity 38
Analysis of the Raven Test 41
  Analytic Strategy 43
  Perceptual Factor 43
  "Copy Strategy" 43
Results 44
Effect of Training on Raven 49
Effect of Training on Raven Items Grouped According to M-demand 53
Effects of Training According to Processing Capacity . 58
Raven Total by M-Level 58
Group Comparison with Anglo Control Students. 60
  Test for Homogeneity of Regression 61
  ANCOVA Results 62
Group Comparisons on Theoretically Defined Raven Subscales 62
  Homogeneity of Regression 62
  ANCOVA 62

IV. RESULTS: CULTURE-LOADING 68

Examination of the Culture-Loaded Hypothesis 68

Identification of Culture-Loaded Test Items 68

Procedure 68

Research Questions 70

Results 71

Validation of the Culture-Loading Hypothesis. 73

Outcomes Consistent with a Culture-Loaded Interpretation. 73
Item Analysis for Culture-Loading 76
  Blacks. 79
  Hispanics 80
Summary 81
Analysis of Culture-Loading on Items Grouped According to M-demand 83
  Hispanics 88
Summary of Results of Culture-Loading 92

Conclusion 96

References 99

Appendix A 106
Appendix B 129

LIST OF TABLES

Table

1 Average Age and Number of Subjects by School, Sex and Race in Training and Control Groups 14

2 Summary of Training Activities and Skill of Problem Area Addressed 21

3 Average Test Scores and Standard Deviations for Training and Control Group on the CEFT, WLT, CCS, FIT and Raven by Group 29

4 Multiple Comparisons Between Treatment and Control Groups on Pre-Training Tests for Blacks, Hispanics and Anglos. . 30

5 Pearson Correlation Coefficients among Age, CEFT, WLT, CCS, FIT and Raven Tests for Black, Hispanic and Anglo Control Students 33

6 Principle-Factor Solution, Factor Analysis with Varimax Rotation of CEFT, WLT, CCS, FIT, Raven and Age for Black, Hispanic and Anglo Control Groups 36

7 M-Level Distribution by Treatment Group for Black, Hispanic and Anglo Groups 40

8 Kruskal-Wallis a-Priori Contrasts between Selected Pairwise Groups on M-Level 42

9 Item Difficulties (Percent Passing) on Raven Test Items for Black, Hispanic, and Anglo Control and Training Groups. . 45

10 Average Score and Standard Deviation of Black, Hispanic and Anglo Students with an M-Level of at Least Zero on the Raven Total and Subscales 50

11 Planned Contrasts for Selected Group Comparisons on Raven Total Score 51

12 Planned Contrasts for Selected Group Comparisons on Raven Subscales 52

13 Grouping of Raven Items According to M-Demand 54

14 Average Score and Standard Deviation of Raven Total and of Items Grouped According to M-Demand of Subjects with M-Level of Zero or Greater 55

15 Planned Contrasts for Selected Group Comparisons on Raven Items Grouped According to M-Demand 56

Table

16 Average Raven Score of Control and Training Subjects by M-Level 59

17 Summary ANCOVA for Comparisons fo Study Groups with Anglo Control Group Subjects on Raven Test with M-Levl as Covariate 63

18 Homogeneity Test for the Regression of Raven M-Demand Subscales onto M-Level in Black, Hispanic and Anglo Training and Control Groups 64

19 Summary ANCOVA of Study Groups with Anglo Control Group on Raven M-Demand Subscales: M-Level as Covariate. . . 65

20 Culture-Loaded Raven Test Items for Black and Hispanic Students by M-Demand 72

21 Summary of Culture-Loaded Outcomes for Specific Group Comparisons 75

22 Comparison on Culture-Loaded and Non-Culture Loaded Raven Items for Black and Anglo Training and Control Groups . 77

23 Comparison on Culture-Loaded and Non-Culture Loaded Raven Items for Hispanic and Anglo Training and Control Groups 78

24 Summary Results of Agreement in Identification of Culture-Loaded and Non-Culture Loaded Raven Test Items 82

25 Percent Passing Culture Loaded and Non-Culture Loaded Items for Black and Anglo Training and Control Groups . 84

26 Group Comparisons on Culture Loaded and Non-Culture Loaded Items for Black and Anglo Training and Control Group Students 85

27 Percent Passing Culture-Loaded and Non-Culture Loaded Items for Hispanic and Anglo Training and Control Groups 89

28 Group Comparisons on Culture-Loaded and Non-Culture Loaded Raven Subscales for Hispanic and Anglo Training and Control Groups. 90

LIST OF FIGURES

Figures

1 "Find the one object that is like the model object in color and size." 7

CHAPTER I

Introduction and Theoretical Position

Several researchers have attempted to provide theories of psychometric test performance which would account for group differences. Probably the most widely known psychometrically based theory is Jensen's (1973) Two Level Theory of Mental Abilities. The theory is described by Jensen as distinguishing between abilities involving the capacity to receive, register and store stimulus information for later recall (Level I) and abilities which involve the transformation, manipulation and integration of stimulus information prior to recall (Level II). According to Jensen, tasks which rely on Level I abilities (primarily) involve rote learning, digit span and other types of simple associative learning. Level II abilities, on the other hand, are involved in tasks like the Raven's Progressive Matrices, Block Design of the WISC and other standard intelligence tests. Jensen argues that group differences (especially Black-White) are due more to differences in Level II abilities than to differences in background experiences. He bases this interpretation on the observation that minorities (Black and Mexican-American) in general perform poorer than Anglos on Level II type tests while performing the same or better on Level I tests (Jensen, 1977).

Questions have been raised concerning processes involved in test performance (Das, 1973a; Das, Kirby & Jarman, 1975; Hunt, 1974; Jensen, 1979; Rohwer, 1971; Sternberg, 1978). Hunt (1974), for example, concludes that there are two qualitatively different ways to solve Raven's Progressive Matrix Problems (a test used by Jensen as an example of "Level II" abilities) which use quite dissimilar processes. Hunt reports that even Spearman noted this, but felt that only one of the processes for solution is related to his general intelligence or *g* factor (Spearman & Wynn-Jones, 1951, as reported by Hunt, 1974, p. 154). Hunt further notes that similar scores and similar patterns of correct and incorrect responses can be attained on Set I of the Raven Progressive Matrices, using either process, and thus lead one to believe (via factor analysis) that the test is measuring the same *g* factor under either manner of solution (in contrast to Spearman's thinking).

As a consequence to this and other issues, other researchers have proposed differences in information processing as an explanation for group differences in test performance (e.g., Case, 1975; Das, 1973b; Rohwer, 1971). Das (1973b) proposed an information processing model, described earlier by Luria (1966), based on "simultaneous" and "successive synthesis." Das contrasts his model with Jensen's stating that the levels model does not take into account individual differences in the tendency to employ different processing strategies.

Along similar lines, but involving different processes is a model proposed by Rohwer (1971). Rohwer suggests that group differences can best be explained in terms of an interaction between the nature of the task (i.e., whether it requires the recall or the application of information) and the individual's propensity to utilize either a formal conceptual process (involving the ability to apply a well-defined set of strategies or rules) or an imaginative process (involving the capacity to depart from formalized conventions in a test situation). Rohwer suggests that group differences occur because minority individuals have not had the same opportunity to de-develop elaborative and conceptual processes to the same degree as majority individuals prior to entering school.

Both Das (1973a) and Rohwer (1971) propose models which attempt to take into account individual differences in the propensity to employ various processing strategies. An important distinction, then, between these models and that of Jensen is that they admit to the possibility that performance on a task can be as much a function of the individual's own idiosyncracies (e.g., choice of processing strategy or cognitive style) as it is determined by the nature of the task.

At the same time, many researchers have been skeptical of psychometric conceptions of intelligence because of their failure to be based on any theory of cognitive abilities (e.g., De Avila, Havassy & Pascual-Leone, 1976; De Vries, 1973; Hathaway, 1973; Hunt 1974; Kohlberg & De Vries, (1969). Hunt (1974) states: "It is inadvisable to have a technology for measuring individual differences which stands apart from a theory of cognition" (p.130).

Case (1975) suggests that SES differences are due to differences in executive repertoire of cognitive processes, rather than information processing capacity. Using a neo-Piagetian approach (Pascual-Leone, 1969, 1970), Case states that performance on Piagetian tasks is affected by the subjects 1) repertoire of executive strategies, 2) cognitive style (e.g., Witkin's, 1950, field differentiation construct), and 3) M-space, the amount of information that the individual can coordinate simultaneously.

Case suggests that it is with respect to differences in executive repertoire of strategies available, which are due to experience, that groups differ.

Such an approach represents a more comprehensive consideration of the issue concerning group differences, the extent of which is well expressed in cross-cultural research (e.g., Buss, 1977; Cole, 1975; Cole & Bruner, 1971; Cole & Gay, 1972; Cole, Gay, Glich & Sharp, 1971). For example, in reference to the relationship between "psychological environment" and group differences, Buss explains:

> There may well be cross-culturally invariant processes (as identified via organismic factors) while at the same time there may also be cross-cultural differences in the learning situations (and hence in the environmental dimensions) in which these invariant processes are applied (Buss, 1977, p. 204).

Thus, at the present time it is safe to say that it is simply not known if groups differ in the processes they use to solve a task or whether the use of different processes still means that the same thing is being measured. However, as noted by Jensen (1979), it does appear that a cognitive processing approach may yield more information concerning what is involved in test performance for individuals in general. Additionally, it is likely that such an approach may also shed more light on the issue of group differences in test performance. At the very least, these possibilities need to be explored.

Information Processing Capacity and Task Complexity

The questions raised above need to be taken into account when one considers the theoretical issues involved in test performance (e.g., see Tuddenham, 1972) aside from those involved in explaining group differences. If we are to understand what a test measures, then we should first know what processes are involved in test performance. In this respect, Hunt (1974) suggests that the style of processing one chooses may be associated with the information demands of the task. He thus concludes that we should "look carefully at the information processing demands of an intelligence test before we decide what the test measures" (p. 130).

A similar conclusion can be reached regarding a comment by Jensen (1980) in which he states that if a learning task (and presumably a test item) "is too complex, everyone, regardless of his IQ, flounders and falls back on simpler processes such as trial-and-error and rote associations" (p.28). Jensen (1979) also points out that increases in task complexity are accompanied by increases in g loading. He states: "(I)t is the task's complexity rather than its content per se that is most related to g" (p. 18). Finally, in the same article he makes the following important comment:

> At present, it seems safe to say, we do not have a real theory of g or intelligence, although we do know a good deal about the kinds of tests that are the most g loaded and the fact that the complexity of mental operations called for by a test is related to g (Jensen, 1979, p. 19).

Recently, a model whereby the processing demands of a task can be determined and compared with the processing capacity of the individual has been developed (Pascual-Leone, 1969, 1970; see also Case, 1972, 1974). Pascual-Leone's model consists of a construct which he terms Mental Operator (M). According to him, M represents the magnitude of the individual's central computing space or M-space, and he defines it as the maximum number of schemes that can be coordinated at any one time (Case, 1972). Pascual-Leone (1969) argues that M is the basic organismic variable underlying psychometric intelligence (i.e., Spearman's g).

Several investigators have attempted to apply *M* directly to psychometric measures of intelligence. For example, Bereiter and Scardamalia (1979), compared the Figure Intersection Task (FIT, Pascual-Leone, 1969) with the Raven Progressive Matrices in an Anglo sample and were able to predict average test performance in both directions. They concluded that, for the most part, the FIT and the Raven test were measuring the same general construct.

Finally, Bachelder and Denny (1977a, 1977b) presented a theory of intelligence based on an individual's *span ability*. Span ability is described much the same as Pacual-Leone's *M* construct. Bachelder and Denny are careful to note that the best measures of span ability are those that involve more complex operations as opposed to simple rote abilities, and which do not allow the subject time to activate a cognitive strategy. It is interesting to note the definition of intelligence provided by Bachelder and Denny:

> Intelligence is the total set of individual difference variables that interact with difficulty or complexity. The more complex the task the more intelligent one needs to be to perform the task. When the task is extremely simple, intelligence is not a relevant variable (1977a, p. 128).

Thus, they state that span ability, (like *M*-space) conforms to their definition of intelligence since it interacts with task complexity.

The idea of an individual's information processing capacity as a set measure of intelligence is not new (Pascual-Leone, 1969). What is new is that it is only now being integrated within the framework of psychometric test performance. The fact that researchers are beginning to use this measure, and that it appears regularly in regard to what the best measures of psychometric intelligence have in common, suggests that processing capacity may offer a more interesting and rewarding measure of intellectual ability.

In addition, information processing capacity and task demands are more amenable to experimental manipulation and control (e.g., see Case, 1974, 1975).

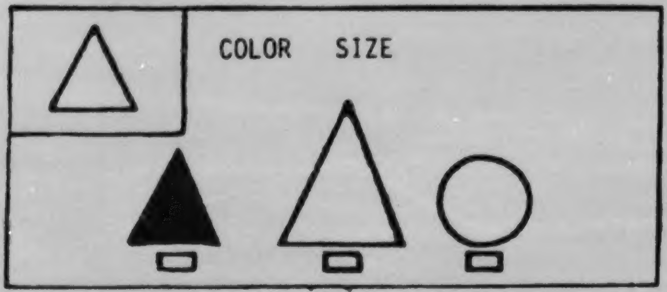
Task Complexity and Culture-Loading

The use of complexity as the common characteristic shared by tests (in varying degrees) offers a different interpretation of test performance than does g or general intelligence. Most important is that complexity is not fixed (as is assumed in classical test theory), but can vary relative to the group or persons attempting the test or item -- that is, it is group specific. For example, the division algorithm is a highly complex task for fourth and fifth grade students. However, for an adult who has over-learned the algorithm and is able to process the task requirements in larger units, it is not as complex.

Similarly, the complexity of many cognitive tasks can vary according to the processing strategy used (e.g., Case, 1975). An example of how a task's complexity can vary is provided in Figure 1.

Figure 1.

"Find the one object that is like the model object in color and size."



The task in Figure 1 is to find which of the three objects is like the model object in terms of color and size. Adults and many children attempt to solve the problem by using a global or simultaneous processing strategy -- i.e., to solve the problem directly by finding the object that is the same color and size. This requires simultaneous processing of both criteria and the distracting cue provided by shape. For most adults, the task does not provide too much difficulty, although some adults will still make mistakes. However, since they cannot process this much information simultaneouly, this strategy often makes the task too complex for young children.

An alternative processing strategy is to employ an analytic or successive processing strategy. In this strategy possible response items are eliminated on the basis of whether they satisfy the first criteria (color). Those remaining are eliminated on the basis of whether they satisfy the second criteria (size). In this way, only one piece of information is processed in each step. Moreover, the distracting cue provided by shape is not even processed at all. Obviously, the second strategy is the preferred one since it reduces the task's complexity to a level which can be solved by most first grade children.

This example illustrates that a task's complexity is not necessarily fixed, unless one can assume that all subjects will be using the same processing strategy. This is why measures of information processing capacity avoid items which allow "chunking" or which can be overlearned through prior experience (e.g., Case 1975; Bachelder & Denny, 1977a). Thus, in the digit-span tests, strings such as 1234 or 1980 are avoided. Bachelder and Denny (1977a) also caution that the rapidity of presentation of items to be recalled should be set at a speed such that no time is allowed for activation of a processing strategy (such as rehearsal) that would reduce the complexity of the task.

Information Processing Capacity and Cognitive Style

Processing strategy can also be influenced by the cognitive style of the subject (Pascual-Leone, 1969; Case & Globerson, 1974). For example, Pascual-Leone (1969) has shown that Witkin's (1950) cognitive style construct of Field Dependence/Independence is related to a subject's tendency to utilize maximum information processing capacity. On tasks which require a complex conceptual response, field-dependent subjects tend not to perform as well as field-independent subjects. Case and Globerson (1974) suggest that the "kind" of field-dependent subject may be an important factor and make a distinction between the subject who is field-dependent because s/he uses little processing capacity and the subject who is field-dependent because s/he is overly sensitive to misleading gestalt-like cues. The latter kind of field-dependence may be the result of unfamiliarity with the elements of the task and hence be attributable to differences in experience.

Group differences in cognitive style have been cited by many researchers (Castañeda, 1976; Laosa, 1978; Laosa & De Avila, 1979; Ramírez & Castañeda, 1974; Ramírez, Castañeda & Herold, 1974; Ramírez & Price-Williams, 1974), and the theory has been used to explain differences in performance on Piagetian Conservation Tasks (Case & Pascual-Leone, 1975), information processing capacity (Case & Globerson, 1974), and other reasoning tasks (Ramírez & Castañeda, 1974). While others have argued that cognitive style differences are in general related to culture (e.g., Ramírez & Castañeda, 1974), research concerning group differences has been equivocal (De Avila & Duncan, 1979). Moreover, Ulibarri and Flemming (1980) reported results which contradict the cultural difference hypothesis and suggest that cognitive style resembles more of an individual trait variable (as opposed to an individual difference variable) in that it tends to be a function of natural experience and simple familiarity with the task situation, indicating a tendency to be overly sensitive to misleading cues, than to a failure to utilize maximum processing capacity. In this sense and for certain tasks, cognitive style is almost a direct measure of culture-loading.

Culture-Loading and Test Bias

The results reported above are especially relevant to the issue of culture-loading and test bias (see Jensen, 1980, Ulibarri, 1982), since, if it is the case that test bias means there is something in a test which makes it easier for one group than for another, then it is possible that this "effect" could be detected whenever differences in processing strategies occur such that differences in the amount of information that must be processed is likely to be affected. If this is the case, it would imply that certain tests are actually more complex for certain students.

This interpretation is consistent with the findings of greater group differences on tests that show the highest _g_-loadings -- i.e., are the most complex (Jensen 1979). If individuals or groups are using different processes to solve a task, and if the task represents different levels of complexity because of this, then different levels of _g_ performance could be affected. That is, so-called _g_ could be a function of the test's complexity as revealed by the group taking the test. In this way, observed group differences on _g_-loaded tests may really reflect differences in task complexity. Thus, if a test is more complex for one group (i.e., requires more information to be processed), then one would expect 1) higher g-loadings and, 2) greater group differences. These results could be due to differences in processing demand for different groups rather than to differences in g-ability.

Stated simply, one interpretation of culture-loading is that a task or test item is culture-loaded if different cognitive processes or processing strategies are likely to be used and if it has the effect of either 1) increasing the number of discrete pieces of information that must be processed by the specific group for whom the test is thought to be biased (e.g., by providing cues which either increase the raw number of pieces of information that must be coordinated, or inhibiting the formation of an executive processing strategy), or 2) decreasing the raw number of discrete pieces of information relative to the group the test is thought to favor (e.g., by providing cues which either reduce the amount of information to be processed, or activating executive processing strategies which aid in coordination of the information to be processed). More specifically, if greater processing demands are required in order for minority subjects to attain the same level of performance on an item as majority subjects, who are equal in processing capacity, then the item or task would be said to be culture-loaded.

In the following, we will describe and discuss a study designed to test the above loosely stated hypothesis that group difference can, to a limited extent, be explained by differences in the way in which children from different racial/ethnic groups approach or solve tests differing in levels of complexity.

CHAPTER II

Design of the Study

Methodology

The basic methodology for this study involves a comparison of the relative contribution of four cognitive/developmental measures of performance on a standard criterion measure of "analytic intelligence and to compare this relative contribution between groups of subjects who received training designed to provide the necessary executive repertoire, relevant to performance on the criterion measure with subjects who did not receive such training.

The basic hypothesis of the study is that children differ in the likelihood of applying the desired cognitive processing strategy, and that when this factor is controlled (through training), performance on the criterion measure is likely to be more similar across ethnic groups. Additionally, it is hypothesized, that such differences result in cultural bias to the extent that different processes are being measured. Stated in another way, cultural-loading on a test is said to occur whenever a test is measuring different aspects of performance. That is, when the assumption that all children taking the test are applying the same processing strategies is not met, that the test is not measuring the same thing in each group.

Subjects

The study was conducted in the Northern California Bay area. Subjects for the study consisted of 134 Black, 83 Hispanic and 74 Anglo children in the fourth and fifth grades. In general, each school in the particular district tends to be composed of one ethnic group. Thus, in order to obtain adequate samples of all three ethnic groups, it was necessary to involve four schools in the study. Additionally, the schools tend to be located in different parts and hence different socio-economic segments of the city. For example, Hispanic students tend to be concentrated in schools in or near the flatland regions of the city, Black students tend to reside nearer the hills, and Anglo students are concentrated more in the eastern hills of the city. Busing provides an additional dimension to the diversity of ethnic make-up in the schools. Nevertheless, in most cases, students of different ethnic groups were not from the same school. In addition, the schools and the student populations are diverse on too many other dimensions besides ethnicity to consider the group comparable per se. Thus, it was not possible to match the backgrounds of the students so that a direct between group analysis would be interpretable.

Table 1 shows the number of subjects and average age by school, sex, and race in training and control groups. The number of children in the training group was determined on the basis of the number of trainers available and other logistic constraints present in each school. Selection of the students for the training group was based on random assignment.

TABLE 1

Average Age[a] and Number of Subjects by School, Sex and Race in Training and Control Groups

School		Black		Hispanic		Anglo	
		Male	Female	Male	Female	Male	Female
CONTROL							
1	Age	122.5	121.9	132.7	126.4	126.0	113.0
	SD	6.94	6.31	10.69	6.93	-0-	-0-
	N	12	22	6	9	1	1
2	Age	128.0	126.3	115.5	--	125.7	125.7
	SD	6.42	3.79	4.95	--	7.48	3.20
	N	6	4	2	-0-	14	4
3	Age	126.9	128.1	123.9	123.8	120.5	--
	SD	10.3	7.4	7.38	11.88	3.54	--
	N	8	7	9	15	2	-0-
4	Age	120.3	120.4	125.7	125.7	127.27	123.7
	SD	3.98	6.37	8.31	11.37	8.56	9.95
	N	6	8	6	3	11	7
TOTAL	Age	123.6		125.5		125.7	
	SD	6.92		9.32		7.78	
	N	73		50		40	
TRAINING							
1	Age	133.0	122.1	122.0	125.3	--	127.0
	SD	12.16	5.63	5.66	7.78	--	5.66
	N	3	13	4	10	-0-	2
2	Age	129.3	125.3	--	131.0	120.3	128.2
	SD	11.04	8.63	--	-0-	9.01	8.23
	N	6	9	-0-	1	7	6
3	Age	124.4	122.7	128.2	125.7	130.3	--
	SD	11.72	8.60	7.48	0.58	8.50	--
	N	9	9	12	3	3	-0-
4	Age	120.5	122.1	--	122.5	126.2	124.1
	SD	7.32	6.13	--	12.02	8.64	9.86
	N	4	8	-0-	2	5	10
TOTAL	Age	124.2		126.2		125.1	
	SD	8.73		7.46		8.99	
	N	61		33		34	

[a] Age in months.

Procedure

The procedure for conducting this study involved identifying Black, Hispanic and Anglo students in the fourth and fifth grades who would volunteer to participate in the study and for whom parent permission was obtained (approximately 85 to 95%). Four schools in a Northern California Bay Area School district were identified on the basis of ethnic composition. Students within each school were assigned to either a training or control group through the use of a random numbers table.

In the early spring of 1980 all subjects were administered four tests: a measure of information processing capacity, a measure of cognitive style, a measure of sensitivity to salient but misleading stimuli, and a neo-Piagetian measure of intellectual development. With the exception of the measure of cognitive style, all tests were administered in a group situation consisting of approximately 15 students per test administrator. The testing sessions lasted approximately one hour each. The information processing test and the measure of sensitivity to perceptual pull were administered together. The information processing test was administered first, followed by the measure of sensitivity to perceptual pull. The neo-Piagetian developmental test was generally administered in a separate group session. In some cases however, the test was administered with the two measures described above. The individually administered measure of cognitive style was administered after all other tests were completed.

Make-up tests were conducted for all students absent during the regular testing schedule. Test administrators had no knowledge whether a student belonged to the training or control group during any of the testing.

Following the initial testing period, students in the training group participated in eight one-hour training sessions conducted over a two week period. Again, make-up was provided so that all training group students completed the eight sessions.

Approximately one month after the initial pre-training tests, all subjects were tested on the Ravens Standard Progressive Matrices according to- the test publisher recommendation. Students in the training and control groups were tested in the same sessions.

For the Hispanic group, Spanish translation was provided for each of the tests and during the training sessions. The following will describe the tests and the training procedure.

Criterion Tests

Information Processing Capacity. The Figure Intersection test (FIT) was used as the criterion measure of processing capacity (Bachelder & Denny, 1977b; Pascual-Leone 1969). In the FIT, students are provided with training on how to take the test. That is, to find the intersection of various overlapping shapes. During the pre-training children are taught first that size, orientation and juxtaposition are irrelevant factors; shape is the only relevant dimension. Second, they are taught to put a dot in each shape that appears on the top-half of the page and then to put one dot where the same shapes are shown overlapping on the bottom of the page.

The FIT consists of seven subscales ranging from two to eight shapes. It has been shown that a subjects' ability to find the intersection is limited according to the number of shapes but increases linearly with age (e.g., De Avila & Havassy, 1974; Pascual-Leone, 1969; Ulibarri, 1974).

A Guttman analysis for the group in this study yielded a coefficient of reproducibility greater than .90 for all groups as well reliabilities (alpha) from .91 to .94.

Cognitive Style. The Children's Embedded Figures Test (CEFT) was adapted by Karp and Korstadt (1963, in Witkin, Oltman, Raskin & Karp, 1971) as a measure of perceptual disembedding. The test requires students to locate a previously seen simple standard figure within a larger complex figure. A score is determined by the number of first correct choices made. Higher scores represent greater field independence. The task requires the subjects to overcome misleading cues provided by the larger, more complex figure. The more independent a subject is from the background or field provided by the larger figure, the more field independent the subject is said to be (Duncan & De Avila, 1979).

Sensitivity to Misleading Cues. The water level test (WLT) (Pascual Leone, 1969, 1970) is a neo-Piagetian measure of cognitive development (Piaget & Inhelder, 1948). Pascual-Leone (1970) has shown the task to be highly related to both information processing capacity and cognitive style. The test consists of a series of illustrated bottles against a three-dimensional rectangular background. Subjects are told to pretend that a picture bottle is half-full, to draw a line showing where the water-level would be and to mark an "x" where the water would be in the bottle. The test contains three subscales consisting of two-dimentional vertical (right-side up and up-side down) bottles and tilted bottles, and three dimentional vertical and tilted bottles. A subjects' score is determined according to deviations from the correct water-level line and correct placement of the location of the water in the bottle. The test is reported in De Avila, et al. (1976), Ulibarri, (1974), and Pascual-Leone, (1972).

Developmental Level. The Cartoon Conservation Scales (CCS) (De Avila, 1977) is neo-Piagetian measure of intellectual development devised from Piagetian theory.

The CCS is made up of eight subtests consisting of 4 items each for a total test length of 32 items. Each item of a particular subtest measures the same concept, each in a slightly different way, by picturing different materials. The eight subtests are listed below in order of increasing difficulty.

1. Conservation of Length
2. Egocentricity/Perspective
3. Conservation of Number
4. Horizontality of Water
5. Conservation of Substance
6. Conservation of Volume
7. Conservation of Distance
8. Probability

The Cartoon Conservation Scales consists of a cartoon-like layout, with the problem presented in three frames on the upper portion of the page and three alternative answer frames located on the lower half of the page. In the first frame of the problem set, an equality or inequality is established. In the second frame, an identity transformation takes place, and in the third frame, a question of equivalence or inequivalence is posed. Three possible answers are based upon the most frequent incorrect responses given by children of this age group. The position of the correct response is varied across the cartoons. Also, the correct response was varied between "yes" and "no" in order to minimize the effects of "yea sayings." Different content was used for each presentation of a concept (e.g., in the conservation of substance, one item used clay as the material and another used beans).

Strictly speaking, the Egocentricity/Perspective scale items are not conservation tasks. Nevertheless, previous research has shown egocentricity items are excellent predictors of conservation and early formal operations. In an egocentricity/perspective item, children are asked to deal with the problem of shifting perspective or point-of-view as represented in three dimensional space.

Analytic Intelligence

The Raven's Standard Progressive Matrices was the criterion measure of analytic ability. The test has been used extensively in the literature and so is only briefly described here. According to Jensen (1974a) the Raven test is a relatively culture-free test. The test consists of 5 subscales of 12 items each. The task is to identify the missing element out of a possible six or eight alternatives. Each item consists of a pattern or sequence of figures. The subject must determine (i.e., abstract) a general rule which, when applied, will lead to selection of the correct response from the possible alternatives. Bachelder and Denny (1979) have shown this test to be highly correlated with the FIT (about .71).

Training

The following is a brief overview of the training procedures. Following this, a more detailed description of each of the training exercises is presented. The purpose of the training was to provide the children with the required executive (i.e., cognitive) repertoire of experiences necessary to perform on the Raven's progressive materials (e.g., see Feurenstein, 1979). This is analogous to what De Avila and Havassy (1974) termed experimental repertoire control and what Spearman and Wynn-Jones (1951, in Hunt, 1974) termed fundaments, that is, controlling for factors considered relevant to taking a test. Without such control, there is a question of whether the test is being administered fairly, or to put it another way, whether the test is likely to measure the same thing for all children taking the test. Generally, the requirement that all children be engaged in the test in the same way, and that they have had equal exposure and opportunity to learn the prerequisite for taking a test, is assumed.

Ten test administrators and eight trainers were used for the study. Administrators and trainers received training on how to administer each of the tests, and on how to conduct the training. The training lasted approximately three full days and consisted of practice taking and administering each of the tests. All test administrators and trainers were college students and some held graduate level degree, all but one had prior experience testing young children.

The training consists of 12 paper and pencil exercises (see Appendix A) administered over an eight day period. The training varied in time from one-half hour to one full hour. The size of the training groups varied from 8 to 12 students per trainer. In each case children were required to work on each exercise until it was completed. Only then, were they allowed to continue on to the next exercise. Some children moved faster than others, but in no case was a child dropped from the training. Children completing the exercises with little difficulty were simply excused while additional help was provided to others. The exercises are summarized in Table 2 together with the day-to-day schedule. Following Table 2 is a more detailed description of the training exercises. The training exercises are based primarily on the work of Feurenstein (1979) and borrow heavily from his research and training.

Day 1 (Pre-Training): The implementation of the exercises began with a review of two of the pre-training tests: the CEFT and the WLT. While this activity was considered beneficial to the overall goals of the training, its basic purpose was to get to know the children and to point out the

TABLE 2

Summary of Training Activities and Skill or Problem Area Addressed

Day	Exercise	Skill or Problem Area Addressed
1	Review of CEFT and WLT.	Demonstration of errors and effect of misleading cues.
2&3	1&2 - Mediated Learning: Subject must find the object that is the same as the model object according to the criterion given (Subjects proceed when they pass criterion test). Exercises 2 and 3 are the same but increase in complexity (i.e., criteria).	Ability to categorize, gathering information from two sources, applying analytic processing strategies, focusing on task instructions, defining problem, ignoring irrelevant but saliant visual information and inhibiting impulsive behavior.
4	3&4 - Dots Training Sheet: Subject must connect seven dots to complete a square and triangle shape.	Practice in visual transport, forming visual structures, using planned behavior, organizing information, gathering precise data, overcoming distracting cues, and forming wholes from parts.
5	5&6 - Figure Completion: Subject must find the part that is missing and complete the figure. Pattern Completion: Subject must complete a pattern to look identical to a model pattern.	Practice in visual transport, completing patterns, and paying attention to detail. Acuity in visual perception, comparative behavior, and pattern recognition.
6	7&8 - Combining Patterns: Subject must combine patterns in either an additive or subtractive manner.	Visual transport (more complex), combining pattern features, acuity in visual perception. Abstracting relationships, applying relationships.
7	9 - Analogies: Subject must abstract the relationship and apply it to complete the matrix. Analogies criterion test.	Transfer of learning to unique problems.
8	10&11&12 - Two by Three Analogies: Three by Three Analogies (Matrices). Matrices criterion test.	Abstracting relationships from two sources of information in two directions, applying analytic processing. Transfer of training to novel problems.

types of errors and answers given by the children. The CEFT and WLT tests were chosen because of the nature of the tests in terms of providing misleading but salient cues. An actual bottle half-filled with water was also used as a demonstration of the WLT task.

During this pre-training period children were asked to provide their own answers and to discuss them with each other. For the WLT task, children were asked to go to the blackboard and draw their solution (i.e., the water level line and location of the water in the bottle). Alternative solutions were also asked for until about three or four different solutions were obtained. A tally was then made to see which solution was preferred. The actual bottle provided the correct answer, to the surprise of many children and one adult observer.

Day 2 & 3 (Mediated Learning): This was by far the most extensive part of the training and will be explained in some detail so that the reader can get a flavor for the training. The first part of training consisted of an exercise conducted by the adult trainer in interaction with the children.

Mediated learning means that there is an interaction between the trainer and the student. That is, the child actively participates in the training. The trainer merely acts as an "adult" mediator who is there to provide direction, place emphasis on certain features and point out errors and correct responses. Thus, the trainer must see that each child performs the task in the context of a discussion on how to solve it. In each task the child first attempts to solve the problem, then it is solved by the group. Incorrect responses are crossed out and correct responses made. The training is designed to develop in the child the appropriate experiences for dealing with the task.

The first exercise consists of 25 cards (see Appendix B). Each card has a model figure in the upper left corner. Under the model figure is a particular dimension label such as "color", "shape", "size" and "pattern." The task is to match or find the object that is most like the model according to the dimension (i.e., criterion) given. Each child is checked to see that the task has been successfully completed before proceeding to the next card. The cards are in the following order: color (items), shape, size and pattern. (Extra cards are available for each criteria in case a child has difficulty.) The children are told the directions and then asked to name the dimension or criteria that is being looked for. The children then mark a "+" sign in the box next to the one they think is correct. The trainer then goes through each response eliciting from the child why it is correct or incorrect. After all the cards are complete, the children are given a criterion mastery test to check for transfer. It consists of three items from each dimension for a total of 12 items. When the children are able to pass all twelve, they go to exercise 2. The following is an example of the dialogue provided to the trainers:

> Directions: Say to the children, "I am going to pass out some booklets that are full of pictures of figures (shapes). In the top left corner (pointing) you will see a figure (triangle) and a word printed below it (color). The game is to find the figure from below that is like the figure in the corner. The word tells you how the figures should be the same. The rule for the game is to use the word to find the figure that is like the one in the corner. When you find the figure (shape) you should put a "plus" in the box below it. O.K., lets try one. Remember, look at the corner figure, and the word, then find the one that is the same according to what the word says. Mark a plus in the box under the one you choose. O.K., the first one says color. We must find one from the bottom that is like the corner figure in "color". What is the color of the top figure? Right it's white. So what are we looking for? (If the children say "white triangle", correct them by asking if the word says triangle or just color. Emphasize that color is the rule (criteria) and that shape doesn't count; only what is given in the "word"

counts.) Have the children mark an answer, then proceed as follows.

O.K., is the first one (point) the one we are looking for? (pause) No. Why? Right, because it's the wrong color (if a child says it is correct because it has color, point out that white is also a color and we are looking for something white). Next, say: what about the middle one? Right, it's the same color. But, before you make a "plus" we should check the last one just in case. Right, it's not the same color so it must be the middle one. So everyone put a "+" in the box below the middle figure.

Following the first exercise children are given a criterion test consisting of 12 items similar to those provided in the training. When children have completed the criterion test without error, they are given exercise 2. This exercise differs in that two dimensions (e.g., color and shape) are given as criteria. When this is completed without errors then the children move to exercise 3.

Day 4 (Dots training sheet): (from Feurenstein, 1979) The dots training sheet consists of a pre-training part and one exercise. In the pre-training children are first shown how to connect four dots to make a square and three dots to make a triangle. Next, the four dots and three dots are juxtaposed in the same picture frame and gradually shown close together in subsequent frames until they overlap. The dots forming the square are at first larger than those forming the triangle. By the last row of frames they are the same size.

The difficulty of the task, of course, increases as the dots become the same size and as the dots forming the square and those forming the triangle overlap. Following training, children are given a mastery test consisting of 19 smaller dotted frames. For some children the task was merely a challenge, while for others it was extremely difficult.

The purpose of the dots task is, in part, to provide fun on an initially easy task, and according to Feurenstein (1979) to provide the children with experience in visual transport forming visual structures, organizing information, gathering precise data, overcoming distracting cues and forming wholes from parts. However, criterion mastery of this task was not required.

Day 5 (Figure and pattern completion): These tasks can be found in Feurenstein (1979). They differ only in that additional figures are involved. In both tasks the problem is to complete a model (criterion) figure. The difference is that in figure completion, a partially completed model figure (e.g., spare, circle, star) is provided together with alternative "parts" of which only one completes the figure. The child must select the correct part. In the pattern completion task, a more complex model figure is shown together with a partially completed figure. The task is to draw in the missing parts so that the partially completed figure is similar to the model or criterion figure. The figure completion task contains patterns that are found in the Ravens test.

Again, according to Feurenstein, these exercises provide practice in visual transport, figural and pattern completion, paying attention to detail, acuity in visual perception and comparative behavior. Subjects completed 12 items on the figure completion and 8 items on the pattern completion tasks. If errors were made, they were pointed out, and the student asked to do them over.

Day 6 (Combining patterns and analogies): The purpose of this day's training was to provide children with experience in combining and subtracting patterns in order to obtain a new pattern. This skill or strategy is then applied to solving figure analogy problems.

The combining patterns task consists of two types of items (4 each). The first involves visually adding two patterns (i.e., overlapping) and selecting from four alternatives the one that would result. The second involves determining what pattern would remain if part of the pattern were removed.

In the visual analogies task the child is presented with a 2 x 2 matrix in which a figure is missing. The task is to select the missing figure out of six alternatives. The child must "abstract" a relationship from the three figures and apply it to one of the alternatives in order to select the correct answer. There are three patterns (i.e., relationships) consisting of four items each for a total of 12 items.

Day 7 (Analogies Criterion Task): This task is simply a more complex version of the previous analogies task. Eight items are given, each of which involves different patterns and somewhat different relationships. Subjects complete this task until reaching 100% mastery. It is the only task given this day and individual help is provided. Aside from the dots exercise, this was the first really difficult task.

Day 8 (2 x 3 and 3 way analogies, and matrices criterion task): Two by three analogies simply involve an extra pattern in the first row of 2 x 2 analogy. However, the children are asked to "draw" the correct answer rather than to select from alternatives. The task appeared easier than 2 x 2 described above. There are 12 items in the task.

The 3 x 3 analogies or matrices problem consists of four sets of items with six alternatives for each set. The task is to select one of the six alternatives in order to complete the matrix.

In the matrices criterion task, there are eight items. The child is asked to draw the correct solution that will complete each matrix. In all of these tasks children are required to attain at least 80% mastery and are provided help (i.e., hints) in order to derive the correct solution.

CHAPTER III

RESULTS: EFFECTS OF TRAINING

Test Performance and Effects of Training

Results of the effects of training presented in this chapter are organized into four sections. The first section examines the relationships between the various tests for Control group students. The second section concerns test performance of both Training and Control group students on the Figure Intersection Test (FIT) as a measure of information processing capacity (i.e., M-level). Following the criterion set by Bereiter and Scardamalia (1979) results for students who achieve an M-level of zero or greater is presented together with an analysis of group comparisons on M-level.

The third section presents results and analysis of training effects for all students obtaining a minimum processing capacity of zero. In this section item difficulties are presented and training effects are examined for the Raven total score, Raven subscales (i.e., published Raven "sets"), and theoretical subscales constructed by grouping items of the same processing requirements (i.e., M-demand).

The fourth section focuses on the effects of training when processing capacity is taken into account. In this section only subjects whose processing capacity is equal to or greater than the processing demands of the items are examined. Results are presented for the Raven total score and the theoretical subscales.

Results for students obtaining an M-level greater than or equal to zero and for those matched with the processing demands of the test are presented so that a complete picture of test performance and the effects of training is obtained. However, since subjects should have the minimum processing capacity for training to be effective in the first place (Case, 1974), matching subjects' processing capacity to processing demands is the main focus of this study.

Relationship Among Tests

The results of the pre-training tests for the training and control groups are given in Table 3. The pre-training tests include the Childrens Embedded Figures Test (CEFT), the Water Level Test (WLT), the Figure Intersection Test (FIT), and the Cartoon Conservation Scales (CCS). Observation of the results indicate that there is little difference between Training and Control Groups for Blacks and a slight difference on the CEFT for the Hispanics. For the Anglo sample there is an apparent trend in favor of the control group. In order to examine performance on the pre-tests, a post hoc analysis was performed using multiple t-test confidence intervals with the Type I error controlled by dividing the alpha level across the four comparisons in each race (i.e., .05/4 .01). The computations for the confidence intervals were performed according to Marascuilo (1971, p. 323). The results for the one Hispanic and the four Anglo comparisons are summarized in Table 4. All other means reported in Table 3 are of such small magnitude (i.e., of no educational significance), as to not warrant testing.

TABLE 3
Average Test Scores and Standard Deviations for Training and Control Group on the CEFT WLT, CCS, FIT and Raven by Group

		Black		Hispanic		Anglo	
		Train	Control	Train	Control	Train	Control
CEFT	$\overline{X}$	18.9	19.3	17.6	19.7	20.4	22.4
	SD	4.24	4.21	5.67	5.85	4.25	3.75
WLT	$\overline{X}$	9.1	8.4	11.7	11.0	11.4	13.7
	SD	4.05	4.04	4.41	4.49	4.43	4.07
CCS	$\overline{X}$	19.4	19.3	19.5	19.6	21.4	23.3
	SD	4.91	5.43	4.02	5.02	6.63	4.77
FIT	$\overline{X}$	17.1	16.8	19.5	20.8	22.1	23.5
	SD	7.13	9.91	9.79	9.65	9.28	8.36
Raven	$\overline{X}$	35.6	28.2	34.2	30.0	41.4	37.4
	SD	10.22	10.37	12.13	10.50	6.56	9.21
	N	61	73	33	50	34	40

TABLE 4
Multiple Comparisons Between Treatment and Control Groups on Pre-Training Tests for Blacks, Hispanics and Anglos

Comparison (Control-Treatment)	$\hat{\psi}$	$SE_{\hat{\psi}}$ [a]	LL	UL
Hispanic (df=81)				
CEFT	2.1	1.2963	-1.32	5.52
Anglo (df=72)				
CEFT	2.0	.93001	-.464	4.46
WLT	2.3	.98876	-.320	4.92
CCS	1.9	1.32922	-1.62	5.42
FIT	1.4	2.05123	-4.04	6.84

$$^{a}SE_{\hat{\psi}} = \sqrt{\left(\frac{S_\rho^2}{N_1} + \frac{S_\rho^2}{N_2}\right)^{\frac{1}{2}}}, \quad S_\rho^2 = \frac{(N_1-1)S^2{}_1 + (N_2-1)S^2{}_2}{N_1 + N_2 - 2}$$

τ determined for $\frac{\alpha}{4} < .01$, two tailed. $df = N_1 + N_2 - 2$

CI $= \hat{\psi} \pm \tau(SE_{\hat{\psi}})$: for Hispanics the critical value for $\tau = 2.64$ and for Anglos $\tau = 2.65$

(Marascuilo, 1971, p. 323)

The results shown in Table 4 indicate that there are no significant differences between Training and Control groups on any of the tests compared. Thus, the randomization procedure for selecting treatment and control students was effective. Nevertheless, the consistent trend demonstrated in the Anglo groups should serve as a caveat in later discussions.

In the following an examination is made of the relationship between the pre-training tests and the Raven. This includes a comparison of the pattern of correlations among the tests in the control groups for each race.

Intercorrelations Among Tests

Previous research (e.g., Bereiter and Scardamalia, 1979; Case & Globerson, 1974) indicates that performance on cognitive style, cognitive developmental and analytic intelligence measures are related to a subject's tendancy to use a large central computing space (M-space) in approaching a cognitive task.

Specifically Bereiter and Scardamalia (1979) report a Pearson correlation of .71 between one version of the FIT and the Raven Progressive matrices in an Anglo sample. They conclude that the Raven and the FIT are essentially measuring the same construct i.e., information processing capacity.

Similarly, Case and Globerson (1974), present empirical evidence in support of the notion that disembedding situations (i.e., CEFT) require a relatively large amount of central computing space in order to solve the task. According to Case and Globerson, a moderate correlation is to be expected between measures such as the CEFT, Raven and FIT tests.

De Avila and Havassy (1974) and Pascual-Leone (1969) demonstrated that performance on neo-Piagetian developmental measures is related to both information processing capacity and cognitive style. In particular, Pascual-Leone argues that a substantial proportion of the variance on developmental and cognitive style measures is due to their shared variance with information processing capacity.

Given this information one would expect a pattern or correlations in which at least a moderate relationship would be exhibited between all of the measures used in this study. In particular, however, a fairly strong correlation would be expected between the FIT and the Raven.

Since a major hypothesis of this study is that culture-loading may occur whenever a test is not measuring the same underlying construct in diverse groups, and that this is a source of bias in test performance, it would be interesting to examine the interrelationship between the tests. If the Raven test exhibits a cultural-bias (i.e., is culture-loaded) then one would expect that the pattern of correlations would not be the same for diverse ethnic groups.

The Pearson correlation matrix for Black, Hispanic and Anglo control group subjects is shown in Table 5. While it is recognized that the three groups are not considered comparably, it should be pointed out that whatever differences existing between the groups is manifest in the pattern of correlations and reflects each group's characteristics as they normally exist in public schools. As such, a comparison of the pattern of correlations is meaningful to the extent that it reflects such differences.

The correlations in Table 5 indicate that the relationship between the tests is similar in that all correlations are significant. However, correlations with age are significant in only the Anglo group. Moreover, the pattern of correlations, as well as the magnitude, is somewhat different. In particular, the correlation between the FIT and Raven for the Anglo

Table 5

Pearson Correlation Coefficients among Age, CEFT, WLT, CCS, FIT and Raven Tests for Black, Hispanic and Anglo Control Students

CONTROL GROUP STUDENTS

	AGE	CEFT	WLT	CCS	FIT
Black (n=73, if $r \geq .19$, $p < .05$)					
CEFT	.12				
WLT	.18	.31			
CCS	.05	.20	.33		
FIT	.12	.56	.47	.27	
RAVEN	.12	.43	.43	.56	.48
Hispanic (n=50, if $r \geq .23$, $p < .05$)					
CEFT	.54				
WLT	.03	.27			
CCS	.16	.30	.27		
FIT	.01	.30	.43	.42	
RAVEN	.06	.33	.37	.55	.49
Anglo (n=40, if $r \geq .26$, $p < .05$)					
CEFT	.37				
WLT	.21	.54			
CCS	.40	.58	.66		
FIT	.36	.47	.56	.54	
RAVEN	.32	.60	.57	.57	.71

group is significally different than that for the Black and Hispanic groups ($p<.05$). This result suggests that the Raven (or the FIT) may be measuring something different in the minority groups. This is examined in the following section.

Factor Analysis

One way to examine the pattern of correlations is through factor analysis. By reducing the number of variables to a smaller set, one may examine the interrelationship between the tests and infer the source of the variance accounting for the observed interrelations in the data. As was suggested in the above discussion, it is expected that there is a common source of variance underlying performance on all the tests. In addition, it is hypothesized that there is an additional source of variability not really related to what the test is intended to measure. This additional source, or factor, is hypothesized to contribute to the culture-loading in a test.

Jensen (1980) demonstrated the utility of a factor analytic approach in examinations of test bias. A difference exists here, however, in that a theoretical rationale has been provided which suggests that the test administered would be applicable to a factor analytic approach. That is, there is an underlying communality in test performance due to information processing capacity and an intervening or common extraneous source due to the processing strategy applied. One would expect then that an additional, or culture-loading factor, would emerge for the minority groups and that tests susceptable to this "bias factor" would load appreciably on this factor.

The method of factor analysis applied is the principal factor solution with varimax rotation. This method was selected because it replaces the main diagonal elements of the correlation matrix with communality estimates and thus automatically produces so-called inferred factors. Table 6 shows the results of the factor analysis for control group students in each ethnic group.

Table 6 shows the first principal factor (unrotated) for each ethnic group and the rotated factor matrix. For the Black and Hispanic groups two factors were extracted and rotated. The Anglo group, however, revealed only the first principal component and thus no rotation was necessary. Table 6 also shows the communalities for each variable (i.e., the total proportion of variance in each variable accounted for by the factors).

The mere fact that the number of factors extracted for the minority groups differs from that of the Anglo group indicates that something different is measured in the combined set of variables. The discussion provided above suggests that a common source of the variance in the set of variables is due to information processing capacity. Thus we would expect that at least one of the factors would represent this construct.

It is fairly clear that for the Anglo group the one factor extracted represents information processing capacity. All of the variables except age load heavily on this factor. It is noted, too, that age, which is also correlated with information processing capacity (i.e., it is a developmental variable), has a restricted variance due to the nature of the sample, i.e., fourth and fifth graders. Thus, the main source of variability, then, for the Anglo group is exactly what would be expected from the battery of tests given.

The minority group factor analyses are not so clear in terms of labeling the factors. It was expected that a factor representing processing strategy would emerge. The unrotated first principal component for each of the minority groups differs from that of the Anglo group in that the load-

TABLE 6

Principle-Factor Solution, Factor Analysis with Varimax Rotation of CEFT, WLT, CCS, FIT, Raven and Age for Black, Hispanic and Anglo Control Groups

BLACKS (Number of Factors =2)

Variables	Unrotated Factor	Rotated Factors Factor 1	Rotated Factors Factor 2	Communality H^2
CCS	.53207	.19491	.59304	.389686
WLT	.59115	.41239	.42775	.35804
CEFT	.57317	.60677	.17051	.39724
FIT	.78198	.90583	.14056	.84028
Raven	.78366	.43778	.69558	.67548
Age	.33576	.02296	.49280	.24289

HISPANICS (Number of Factors =2)

	Unrotated 1st Factor	Rotated Factors Factor 1	Rotated Factors Factor 2	Communality H^2
CCS	.70732	.73391	.13858	.55783
WLT	.55202	.46183	.26413	.28306
CEFT	.56588	.36044	.51280	.39288
FIT	.65396	.63773	.20049	.44689
Raven	.71721	.79232	.05516	.63082
Age	.47371	.04870	.87786	.77302

ANGLOS (Number of Factors =1)

	Unrotated 1st Factor	Communality H^2
CCS	.80164	.64263
WLT	.73386	.53854
CEFT	.71676	.51374
FIT	.76227	.58105
Raven	.81300	.66097
Age	.32175	.10352

ings are lower due to the additional factor. Moreover, the communalities indicate that, in general, a greater proportion of the variance in each variable is accounted for by a single factor in the Anglo group than is explained with two factors in the minority group.

The results of the rotated factor matrix show that the Raven test is loaded on both factors in the Black sample but on only one factor in the Hispanic sample. This suggests that the nature of culture-loading hypothesized in this study is not reflected as much with the Hispanic group, at least with the set of variables included in the analysis. Thus, while a second factor emerged, the Raven test did not load appreciably on this factor, and consideration of the results for Blacks and Hispanics must be considered separately.

A possible interpretation of the results for Blacks is that the first rotated factor represents a processing strategy factor while the second represents an analytic ability or processing capacity factor. The rationale for this interpretation is the fact that the FIT and Raven are loaded on different factors. Moreover, the CEFT also loads more heavily on the factor defined primarily by the FIT.

The theoretical discussion provided above suggested that the set of tests have information processing capacity in common. However, the tests also have a cognitive style factor in common. Cognitive style is known to effect performance on information processing tasks and analytic tasks which require a disembedding solution (Case & Globerson, 1974). Thus, the interpretation of the first factor as a processing strategy factor is consistent with this expectation. Additionally, the CCS also loads primarily on the 2nd factor defined by the Raven. The one curious result is the rather moderate loading of the FIT on factors defined by the Raven.

The results for the Hispanics are clearer. For example, the first rotated factor can be identified as an analytic or information processing factor. The second factor is defined by age and the CEFT. The WLT and the FIT load only moderately on this factor. The important thing to note is that while there is an additional factor associated with processing strategy, the Raven did not load on this factor. Nevertheless, cognitive style did not show the same relationship in the battery of tests. This indicates that it is a source of variance between subjects (i.e., age groups) but is not necessarily related to performance on the Raven as was expected.

The factor analyses indicate that the main difference in patterns of correlations is age related. The correlation matrices indicated this and the factor analysis demonstrated it. In general, it appears that the Black and Anglo group differ most in terms of variables related to Raven test performance while the Hispanics are somewhat similar to the Anglo group. Two factors did emerge for the minority groups, and only one for the Anglos.

In the following, the results of the FIT test as a measure of information processing level (M-level) and the effect of training on Raven test performance are presented. Following this is an item analysis of the Raven and an examination of the culture-loading hypothesis.

Information Processing Capacity

Information processing capacity or M-level is defined as the number of discrete pieces of information that can be processed simultaneously. The set measure of M-level is the Figure Intersection Test (FIT, Pascual-Leone, 1969). Scoring for the FIT to obtain a subject's M-Level followed the procedures described by Bereiter and Scardamalia (1979). In this procedure the percentage of correct responses on each FIT subscale are summed and a constant of 1.5 is subtracted from the total. The result is the subject's "M-level."

Some subjects will obtain an M-level less than zero when this procedure is used. Subjects obtaining M-levels less than zero are thought to have done so because of inattentivness or failure to grasp the task instructions (Bereiter & Scardamalia, 1979). Consequently, following the precedent established by Bereiter and Scaradamalia, such subjects are dropped from the analysis. The results of M-level assignments are given in Table 7, together with the number of subjects obtaining an M-level of less than one for each group.

Table 7 shows the number and percentage of students obtaining a given M-level greater than zero on the FIT. The results are roughly equivalent to what would be expected for subjects of this age group (Case, 1972). The mean M-level rank, average M-level, standard deviation and median M-level for Black, Hispanic and Anglo subjects are shown at the bottom of the Table.

In order to test for group differences in the distribution of M-levels, a Kruskal-Wallis one-way analysis of variance on the ranks (following the cell-means procedure described by Marascuilo and Levin (1976) was performed on the ranks with planned comparisons on selected groups. The cell-means model of analysis allows for tests of hypotheses normally associated with either a nested analysis (i.e., between treatments within race) or a fully-crossed analysis (i.e., interaction). The cell-means model is basically a one-way analysis of variance in which each group is treated as a single block. This results in six groups defined as follows: one group each for Black, Hispanic and Anglo Training groups and one each for Black, Hispanic, and Anglo Control groups.

Table 7

M-Level^a Distribution by Treatment Group for Black, Hispanic and Anglo Groups

M-Level	Black				Hispanic				Anglo			
	Train		Control		Train		Control		Train		Control	
	N	%	N	%	N	%	N	%	N	%	N	%
0	6	10.3	13	20.6	3	10.3	1	2.3	1	3.2	1	2.6
1	13	22.4	7	11.1	2	6.9	5	11.4	2	6.5	4	10.5
2	16	27.6	9	14.3	4	13.8	7	15.9	3	9.7	1	2.6
3	12	20.7	13	20.6	6	20.7	6	13.6	7	22.6	7	18.4
4	7	12.1	8	12.7	7	24.1	12	27.3	6	19.4	11	28.9
5	4	6.9	12	19.0	7	24.1	9	20.5	9	29.0	12	31.6
6	0	0	1	1.6	0	0	4	9.1	3	9.1	2	5.3
N	58		63		29		44		31		38	
Mean Rank	169.16		126.24		100.48		150.29		111.88		98.38	
M < 0 N	3		10		4		6		3		2	
Mean	2.41				3.36				3.75			
SD	1.636				1.602				1.499			
Median												
N	121				73				69			

^{a}M-Level = $\Sigma P_i - 1.5$, P_i = Percent items correct on FIT subscales (Bereiter & Scardamalia, 1979)

Nine contrasts were computed for comparisons between Training and Control groups in each ethnic group (3), between Black-Anglo and Hispanic-Anglo within Training and Control groups (4), and two interaction contrasts comparing the differences between the Training and Control groups for Black and Hispanic with the difference between the Anglo Training Control groups. The planned contrasts were computed according to the procedures described in Marascuilo and McSweeney (1977). Since both the full nested analysis and the fully crossed analysis allow for an overall .15 type I error rate (alpha), this error rate was distributed across the nine contrasts using probalities obtained from Dunn's (1961) table of critical values. With this procedure each contrast is tested at an alpha level of .0167. Table 8 presents the results of the Kruskal-Wallis analysis on the ranks.

Of the nine contrasts shown in Table 8, two are significant. These involved the comparisons between Black vs. Anglo Training and Control groups. Direct interpretation is difficult because of the confounding of school attended, socio-economic status and male-female distributions in each ethnic group. In addition, the factor for the Black students in the sample suggests that the FIT test may not be measuring the same thing in each group. The comparisons were performed to examine the distributions of the samples in the study and are not amenable to generalizations beyond this purpose.

Analysis of the Raven Test

There are sixty items in the Raven test. The items are grouped into five subscales of 12 items each. Each item within a subscale becomes progressively more difficult as does each successive subscale. The subscales are also dependent upon different cognitive processes. That is, different

Table 8

Kruskal-Wallis a Priori Contrasts[a] between Selected Pairwise Groups on M-Level

Comparison	$\hat{\Psi}$	$SE_{\hat{\Psi}}$[b]	LL	UL
Ψ_1: Black Training vs. Control	18.869	13.6386	-13.7272	51.4652
Ψ_2: Hispanic Training vs. Control	14.366	17.9266	-28.4786	57.2106
Ψ_3: Anglo Training vs. Control	2.102	18.1391	-41.2385	45.4425
Ψ_4: Black Training vs. Anglo Training	68.671	16.6749	28.8180	108.5240*
Ψ_5: Hispanic Training vs. Anglo Tratining	25.757	19.3624	-20.5191	72.0331
Ψ_6: Black Control vs. Anglo Control	51.618	15.3944	14.8254	88.4106*
Ψ_7: Hispanic Control vs. Anglo Control	13.493	16.5978	-26.1758	53.1618
Ψ_8: $\Psi_1 - \Psi_3$ (interaction)	16.767	22.6948	-37.4336	71.0076
Ψ_9: $\Psi_2 - \Psi_3$ (interaction)	12.264	25.5027	-48.6875	73.2155

[a] Corrected for tie values, @ 2.9708 (Marascuilo & McSweeney, 1977)

[b] $T_{\text{Dunn}} = 2.39$, $Q = 9$, total $\alpha = .15$

* Statistically significant $\alpha = .017$

cognitive processess may effect the complexity of the item in terms of the amount of information that must be processed in order to correctly solve the item. For example, the easier subscales can be solved by a global or visual processing strategy, while later subscales are dependent more on an analytic processing strategy. According to Bereiter and Scardamalia (1979), there are at least three factors which effect the difficulty level of Raven items. The three strategies are summarized in the following:

1. Analytic Strategy: Three types of problems are identified which are a function of whether the item type involves a) pattern repetition, b) elements permuted, or 3) progressions.
2. Perceptual Factor: Influences the difficulty level of an item in an undetermined way; the item may be made easier or more difficult depending on the nature of the perceptual factor and that of the item. Perceptual factors are grouped under a) Gestalt effects and b) embedding of figures.
3. "Copy Strategy:" This is simply a means by which subjects compare consecutive figures to determine the closest match. This strategy was identified on the basis of subjects' eye movements.

The theoretical discussion provided in Chapter I leads to the prediction that training would have a greater effect on the more difficult items (i.e., those involving more complexity) and on those items involving perceptual factors. It is hypothesized that non-trained minority subjects tend to respond more to perceptual factors than do their Anglo counterparts. For complex tasks, a perceptual processing strategy would require more processing capacity on the part of the subject in order to overcome the distraction caused by the misleading but salient perceptual factors. For this reason, it is hypothesized that many students perform poorly on analytic and disembedding processing tasks. The question thus becomes one of removing these differences so that assessment of the desired characteristics can be made. The research questions to be examined in this part of the study are summarized in the following:

H_0: Training will have a positive effect on test performance for each ethnic group; however, the change (i.e., gains) will be greater for minority subjects.

H_0: Training will have its greatest effect on the most difficult subscales of the Raven (difficulty based on percent passing).

H_0: Training will have its greatest effect on the most complex Raven test items (complexity defined in terms of item processing demands).

H_0: Raven test items identified, a-priori, as culture-loaded will show a greater effect due to training than items not identified as culture loaded. That is, training will have the greatest effect on the items identified as culture-loaded.

Results

Percentage passing each Raven item, average subscale and Raven total score are provided in Table 9 for each ethnic group. There are 216 means provided in Table 9. Of initial interest are Raven total test and mean subscale scores. From Table 9 it can be seen that the average difference between all subjects in the training and control groups for Blacks, Hispanics, and Anglos is 7.4, 4.2 and 3.9 repectively for the Raven total. These results represent the performance of all subjects regardless of M-level. The analysis reported below, however, is restricted to performance of subjects obtaining an M-level of at least zero for reasons discussed earlier, so that the results will be consistent with the item analysis

TABLE 9
Item Difficulties (Percent Passing) on Raven Test Items for Black, Hispanic, and Anglo Control and Training Groups

Black Students

	Control			Training	
Item	Percent Passing	Standard Deviation	Item	Percent Passing	Standard Deviation
R1	1.0000	0.0	R1	0.9836	0.1280
R2	1.0000	0.0	R2	1.0000	0.0
R3	0.9726	0.1644	R3	1.0000	0.0
R4	1.0000	0.0	R4	1.0000	0.0
R5	0.9041	0.2963	R5	0.9836	0.1280
R6	0.9041	0.2963	R6	1.0000	0.0
R7	0.6164	0.4896	R7	0.8033	0.4088
R8	0.8219	0.3852	R8	0.8325	0.3576
R9	0.7260	0.4491	R9	0.9016	0.3003
R10	0.5890	9.4954	R10	0.7049	0.4599
R11	0.4110	0.4954	R11	0.5738	0.4986
R12	0.1781	0.3852	R12	0.3934	0.4926
ASUB	9.1370	1.9601	ASUB	10.1803	1.4777
R13	0.9863	0.1170	R13	0.9836	0.1280
R14	0.8904	0.3140	R14	0.9508	0.2180
R15	0.7671	0.4256	R15	0.8197	0.3877
R16	0.6575	0.4778	R16	0.8689	0.3404
R17	0.6027	0.4927	R17	0.6066	0.4926
R18	0.4521	0.5011	R18	0.6066	0.4926
R19	0.4247	0.4977	R19	0.6393	0.4842
R20	0.4110	0.4954	R20	0.6230	0.4887
R21	0.4384	0.4996	R21	0.6885	0.4669
R22	0.5205	0.5030	R22	0.8033	0.4008
R23	0.5342	0.5023	R23	0.6885	0.4669
R24	0.2055	0.4068	R24	0.3770	0.4887
BSUB	6.8767	3.3827	BSUB	8.6557	2.8803
R25	0.7945	0.4068	R25	0.8852	0.3214
R26	0.7123	0.4558	R26	0.7049	0.4599
R27	0.6712	0.4730	R27	0.8033	0.4008
R28	0.5753	0.4977	R28	0.7049	0.4599
R29	0.5479	0.5011	R29	0.7213	0.4521
R30	0.4521	0.5011	R30	0.6393	0.4842
R31	0.3286	0.4730	R31	0.7377	0.4435
R32	0.3151	0.4678	R32	0.5246	0.5035
R33	0.5068	0.5034	R33	0.7541	0.4342
R34	0.2192	0.4166	R34	0.2623	0.4435
R35	0.1370	0.3462	R35	0.1311	0.3404
R36	0.1096	0.3145	R36	0.0328	0.1786
CSUB	5.3699	2.9557	CSUB	6.916	2.7185
R37	0.8082	0.3964	R37	0.9344	0.2496
R38	0.6301	0.4861	R38	0.7705	0.4240
R39	0.5479	0.5011	R39	0.7213	0.4521
R40	0.6164	0.4896	R40	0.7377	0.4435
R41	0.6436	0.4822	R41	0.8361	0.3733
R42	0.5205	0.5030	R42	0.6557	0.4791
R43	0.3699	0.4861	R43	0.4426	0.5008
R44	0.5068	0.5034	R44	0.6230	0.4887
R45	0.3151	0.4678	R45	0.3934	0.4926
R46	0.2877	0.4558	R46	0.5082	0.5041
R47	0.1096	0.3145	R47	0.1211	0.3404
R48	0.0548	0.2292	R48	0.0656	0.2496
DSUB	5.3836	3.2814	DSUB	6.8197	2.8490
R49	0.2740	0.4491	R49	0.5574	0.5008
R50	0.2877	0.4558	R50	0.5246	0.5035
R51	0.2192	0.4166	R51	0.4098	0.4959
R52	0.1370	0.3462	R52	0.2951	0.4599
R53	0.0822	0.2766	R53	0.3443	0.4791
R54	0.0959	0.2965	R54	0.2295	0.4240
R55	0.1233	0.3310	R55	0.1311	0.3404
R56	0.0548	0.2292	R56	0.1639	0.3733
R57	0.0548	0.2292	R57	0.1475	0.3576
R58	0.0548	0.2292	R58	0.0492	0.2180
R59	0.0959	0.2963	R59	0.0328	0.1796
R60	0.0274	0.1644	R60	0.0656	0.2496
ESUB	1.5205	1.4153	ESUB	2.9508	2.4928
Raven	28.1507	10.3664	Raven	35.5738	10.2233
	(N=73)			(N=61)	

TABLE 9 (Continued)
Item Difficulties (Percent Passing) on Raven Test Items for Black, Hispanic, and Anglo Control and Training Groups

Hispanic Students

	Control			Training	
Item	Percent Passing	Standard Deviation	Item	Percent Passing	Standard Deviation
R1	1.0000	0.0	R1	1.0000	0.0
R2	1.0000	0.0	R2	1.0000	0.0
R3	1.0000	0.0	R3	0.9394	0.2423
R4	0.9800	0.1414	R4	0.9394	0.2423
R5	0.9800	0.1414	R5	0.9697	0.1741
R6	0.9400	0.2399	R6	0.9697	0.1741
R7	0.6600	0.4785	R7	0.6970	0.4667
R8	0.7600	0.4314	R8	0.6364	0.4885
R9	0.8400	0.3703	R9	0.9394	0.2423
R10	0.7600	0.4314	R10	0.8788	0.3314
R11	0.6200	0.4903	R11	0.6061	0.4962
R12	0.4200	0.4986	R12	0.3939	0.4962
ASUB	9.9600	1.9268	ASUB	9.9697	2.2289
R13	0.9800	0.1414	R13	1.0000	0.0
R14	0.9600	0.1979	R14	1.0000	0.0
R15	0.7400	0.4431	R15	0.8182	0.3917
R16	0.6800	0.4712	R16	0.6970	0.4667
R17	0.5600	0.5014	R17	0.7576	0.4352
R18	0.5000	0.5051	R18	0.5455	0.5056
R19	0.4200	0.4986	R19	0.4545	0.5056
R20	0.4400	0.5014	R20	0.5455	0.5056
R21	0.5000	0.5051	R21	0.5758	0.5019
R22	0.6200	0.4903	R22	0.6667	0.4787
R23	0.4200	0.4986	R23	0.4848	0.5075
R24	0.2400	0.4314	R24	0.3030	0.4667
BSUB	7.0600	3.0932	BSUB	7.8485	3.4652
R25	0.8600	0.3505	R25	0.8788	0.3314
R26	0.7800	0.4185	R26	0.7273	0.4523
R27	0.7400	0.4431	R27	0.6970	0.4667
R28	0.5600	0.5014	R28	0.5455	0.5056
R29	0.6400	0.4849	R29	0.6970	0.4667
R30	0.4000	0.4949	R30	0.4848	0.5075
R31	0.5600	0.5014	R31	0.7879	0.4151
R32	0.4000	0.4949	R32	0.5152	0.5075
R33	0.4800	0.5047	R33	0.7576	0.4352
R34	0.1600	0.3703	R34	0.3030	0.4667
R35	0.1000	0.3030	R35	0.2727	0.4523
R36	0.1400	0.3505	R36	0.0606	0.2423
CSUB	5.8200	2.8692	CSUB	6.7273	2.8093
R37	0.8400	0.3703	R37	0.9394	0.2423
R38	0.6000	0.4949	R38	0.7273	0.4523
R39	0.5400	0.5035	R39	0.7576	0.4352
R40	0.5600	0.5014	R40	0.5152	0.5075
R41	0.7400	0.4431	R41	0.7576	0.4352
R42	0.5000	0.5051	R42	0.5758	0.5019
R43	0.4000	0.4949	R43	0.6061	0.4962
R44	0.4600	0.5035	R44	0.4545	0.5056
R45	0.2600	0.4431	R45	0.2727	0.4523
R46	0.2000	0.4041	R46	0.4848	0.5075
R47	0.1200	0.3283	R47	0.1515	0.3641
R48	0.0600	0.2399	R48	0.0606	0.2423
DSUB	5.2800	2.9969	DSUB	6.3030	3.1373
R49	0.5000	0.5051	R49	0.5455	0.5056
R50	0.3400	0.4785	R50	0.6061	0.4962
R51	0.2800	0.4536	R51	0.6061	0.4962
R52	0.1800	0.3661	R52	0.4545	0.5056
R53	0.2200	0.4185	R53	0.4242	0.5019
R54	0.1400	0.3505	R54	0.2727	0.4523
R55	0.1000	0.3030	R55	0.0909	0.2919
R56	0.1200	0.3283	R56	0.1818	0.3917
R57	0.0400	0.1979	R57	0.0909	0.2919
R58	0.0600	0.2399	R58	0.0606	0.2423
R59	0.0600	0.2399	R59	0.0303	0.1741
R60	0.0800	0.2740	R60	0.0606	0.2423
ESUB	2.1200	2.0368	ESUB	3.4242	2.6579
Raven	30.0400	10.5016	Raven	34.2424	12.1399
	(N=30)			(N=33)	

TABLE 9 (Continued)
Item Difficulties (Percent Passing) on Raven Test Items for Black, Hispanic, and Anglo Control and Training Groups

Anglo Students

	Control			Training	
Item	Percent Passing	Standard Deviation	Item	Percent Passing	Standard Deviation
R1	1.0000	0.0	R1	1.0000	0.0
R2	1.0000	0.0	R2	1.0000	0.0
R3	0.9750	0.1581	R3	1.0000	0.0
R4	0.9750	0.1581	R4	1.0000	0.0
R5	0.9500	0.2207	R5	1.0000	0.0
R6	0.9750	0.1581	R6	1.0000	0.0
R7	0.8750	0.3349	R7	0.8529	0.3595
R8	0.8750	0.3349	R8	0.9412	0.2388
R9	0.9000	0.3038	R9	1.0000	0.0
R10	0.8000	0.4051	R10	0.8235	0.3870
R11	0.6500	0.4830	R11	0.6765	0.4749
R12	0.4750	0.5057	R12	0.4706	0.5066
ASUB	10.6250	1.5801	ASUB	10.7647	1.1297
R13	1.0000	0.0	R13	0.9706	0.1715
R14	0.9500	0.2207	R14	0.9706	0.1715
R15	0.9500	0.2207	R15	0.8824	0.3270
R16	0.8750	0.3349	R16	0.7941	0.4104
R17	0.8250	0.3848	R17	0.6765	0.4749
R18	0.7750	0.4229	R18	0.6176	0.4933
R19	0.7500	0.4365	R19	0.5882	0.4996
R20	0.5250	0.5057	R20	0.7941	0.4104
R21	0.6500	0.4830	R21	0.7647	0.4306
R22	0.7000	0.4641	R22	0.7941	0.4104
R23	0.6750	0.4743	R23	0.7059	0.4625
R24	0.4250	0.5006	R24	0.4412	0.5040
BSUB	9.0750	2.7955	BSUB	9.0000	3.0050
R25	0.9000	0.3038	R25	0.9706	0.1715
R26	0.9250	0.2667	R26	0.8529	0.3595
R27	0.8000	0.4051	R27	0.9706	0.1715
R28	0.8000	0.4051	R28	0.7941	0.4104
R29	0.8000	0.4051	R29	0.8824	0.3270
R30	0.6750	0.4743	R30	0.7941	0.4104
R31	0.6500	0.4830	R31	0.9118	0.2879
R32	0.6000	0.4961	R32	0.7353	0.4478
R33	0.6500	0.4830	R33	0.9118	0.2879
R34	0.2750	0.4522	R34	0.5294	0.5066
R35	0.1500	0.3616	R35	0.4706	0.5066
R36	0.0750	0.2667	R36	0.0882	0.2879
CSUB	7.2750	2.6697	CSUB	8.9118	1.9285
R37	0.9500	0.2207	R37	0.9412	0.2388
R38	0.8750	0.3349	R38	0.9706	0.1715
R39	0.8000	0.4051	R39	0.9412	0.2388
R40	0.8500	0.3616	R40	0.8529	0.3595
R41	0.7750	0.4229	R41	1.0000	0.0
R42	0.7750	0.4229	R42	0.8529	0.3595
R43	0.6500	0.4830	R43	0.6471	0.4851
R44	0.5750	0.5006	R44	0.7059	0.4625
R45	0.4750	0.5057	R45	0.4118	0.4996
R46	0.5500	0.5038	R46	0.7941	0.4104
R47	0.1500	0.3616	R47	0.2353	0.4306
R48	0.0	0.0	R48	0.0588	0.2388
DSUB	7.4250	2.5809	DSUB	8.4412	1.7441
R49	0.6000	0.4961	R49	0.5588	0.5040
R50	0.5250	0.5057	R50	0.6471	0.4851
R51	0.2250	0.4229	R51	0.5588	0.5040
R52	0.4000	0.4961	R52	0.4706	0.5066
R53	0.2250	0.4229	R53	0.5000	0.5075
R54	0.2250	0.4229	R54	0.3624	0.4933
R55	0.2750	0.4522	R55	0.3529	0.4851
R56	0.1500	0.3616	R56	0.2647	0.4478
R57	0.0	0.0	R57	0.2059	0.4104
R58	0.1000	0.3038	R58	0.2353	0.4306
R59	0.0500	0.2207	R59	0.0	0.0
R60	0.0500	0.2207	R60	0.0882	0.2879
ESUB	3.0250	2,1302	ESUB	4.2647	2.6089
Raven	37.3750	9.2062	Raven	41.3529	6.5592
	(N=40)			(N=34)	

Page 3

of the Raven for culture-loading. The analysis of Raven total test performance is examined. Second, analysis is made of Raven subscale performance according to those defined in the Raven test manual. Finally, analysis is made of theoretically constructed subscales grouped according to item M-demands.

Analysis of this data was performed using the Cell-means model with planned comparisons described by Marascuilo and Levin (1976). As noted above the Cell-means model was selected because it allows for analysis of both the research questions concerning within-group training differences (nested analysis) and a comparison of the gains between groups (interaction analysis) without confounding the Type I error rate. The analysis is essentially a one-way analysis of variance of the six groups in the study. One-tail planned contrasts are then computed for selected comparisons with the total alpha associated with the full factorial design (i.e., treatment by group: 2 x 3) distributed accross the planned comparisons.

The alpha level for a full factorial analysis is .15. This represents .05 for each of the two main-effects of treatment and ethnic group, and .05 for their interaction. For the analysis five contrasts are made. These include three within-group training effects contrasts (i.e., nested analysis of training vs. control within race) and two interaction contrasts comparing training effects for Black and Hispanic groups with the Anglo groups. The within group contrasts were computed with an alpha level of .10, using critical _t_-values such that each contrast is tested at alpha equal to .10/3=.03. The two interaction contrasts were computed with alpha controlled at .05 (the level allowed in a fully crossed analysis) so that each comparison is made at .025. The total experiment-wide alpha is .15 in agreement with a fully crossed two-way analysis of variance. The standard error term is obtained from the one-way analysis.

In the following section the results for the total Raven test are presented first, followed by an analysis of Raven subscales and then the theoretically constructed subscales.

Effect of Training on Raven

Average test performance on the Raven total and subscales for each group are given in Table 10. The results of the Cell-means analysis of the Raven total score is reported in Table 11.

The results shown in Table 11 indicate a significant difference in favor of the treatment groups for Black and Hispanic students, but not for Anglo students. At the same time, however, the significant training effect for minority students was not significantly greater than the difference in the Anglo group (i.e., no interaction effect).

These results do not support the hypothesis that training would be effective for all groups, nor do they support the hypothesis that the training effect would be greater for minority students than for Anglo students. On the other hand, partial support is obtained indirectly in that a significant training effect occurred for minority students but not for Anglo students. Stated in this way, training was effective for minority students but not for Anglo students on the Raven test overall.

The same hypotheses were tested for the Raven Subscales. The results of the analysis is shown in Table 12.

Significant training effects occurred for the Black training group on all subscales with the exception of subscale D. Hispanics showed a significant difference in test performance on subscale E, while the Anglo train-

Table 10

Average Score and Standard Deviation of Black, Hispanic and Anglo Students with an M-Level of at Least Zero on the Raven Total and Subscales

		Black		Hispanic		Anglo	
		Training	Control	Training	Control	Training	Control
Subscale A	$\bar{X}$	10.2	9.2	10.5	10.2	10.8	10.7
	SD	1.50	2.03	1.60	1.83	1.13	1.43
Subscale B	$\bar{X}$	8.7	7.1	8.4	7.4	9.2	9.4
	SD	2.88	3.44	3.18	3.09	2.86	2.52
Subscale C	$\bar{X}$	6.9	5.5	7.0	6.1	9.1	7.6
	SD	2.78	2.99	2.60	2.83	1.78	2.37
Subscale D	$\bar{X}$	6.7	5.8	6.7	5.5	8.7	7.7
	SD	2.89	3.16	3.10	2.86	1.29	2.21
Subscale E	$\bar{X}$	3.0	1.6	3.7	2.2	4.4	3.1
	SD	2.48	1.46	2.70	2.12	2.68	2.15
Raven	$\bar{X}$	35.7	28.9	36.3	31.3	42.3	38.5
	SD	10.34	10.39	11.15	10.46	5.71	10.59
	N	58	63	29	44	31	

Table 11

Planned Contrasts[a] for Selected Group Comparisons on Raven Total Score

Comparison	$\hat{\Psi}$	$SE_{\hat{\Psi}}$	One-tail Confidence Interval[b]
Ψ_1 = Black Training vs. Control	6.67	1.7686	2.86
Ψ_2 = Hispanic Training vs. Control	5.03	2.3247	.0319
Ψ_3 = Anglo Training vs. Control	3.85	2.3522	-1.2072
$\Psi_4 = \Psi_1 - \Psi_3$	2.82	2.9430	-2.9483
$\Psi_5 = \Psi_2 - \Psi_3$	1.18	3.3071	-5.3019

[a] Cell-means analysis (Marascullo & Levin, 1976): $\Psi = \hat{\Psi} - t(SE_{\hat{\Psi}})$

$$SE_{\hat{\Psi}} = MSW \sum_{i=1}^{k} \frac{a_i^2}{N_i}, \quad MSW = 94.4624$$

[b] One-tall $\underline{t}$ critical-value for ψ_1, ψ_2, and ψ_3 equals 2.15; $\alpha_{tot} = .10$, $\alpha_{\hat{\psi}} = .033$ One-tall $\underline{t}$ critical-value for Ψ_4 and Ψ_5 equals 1.96, $\alpha_{tot} = .05$, $\alpha_{\hat{\psi}} = .025$
Values are statistically significant if greater than zero.

Table 12

Planned Contrasts for Selected Group Comparisons on Raven Subscales

Comparison	$\hat{\Psi}$	$SE_{\hat{\Psi}}$	One-tail Confidence Interval[a]
Subscale A			
Ψ_1: Black Train. vs. Control	1.02	.3032	.3681*
Ψ_2: Hispanic Train. vs. Control	.30	.3985	-.5568
Ψ_3: Anglo Train. vs. Control	.10	.4032	-.7670
Ψ_4: $\Psi_1 - \Psi_3$	.92	.5045	-.0688
Ψ_5: $\Psi_2 - \Psi_3$	.20	.5669	-.9112
Subscale B			
Ψ_1: Black Train. vs. Control	1.59	.5538	.3992*
Ψ_2: Hispanic Train. vs. Control	1.0	.7280	-.5651
Ψ_3: Anglo Train. vs. Control	-.14	.7366	-1.7237
Ψ_4: $\Psi_1 - \Psi_3$	1.73	.9216	-.0763
Ψ_5: $\Psi_2 - \Psi_3$	1.14	1.0356	-.8960
Subscale C			
Ψ_1: Black Train. vs. Control	1.42	.4859	.3753*
Ψ_2: Hispanic Train. vs. Control	.90	.6387	-.4732
Ψ_3: Anglo Train. vs. Control	1.55	.6463	.1605*
Ψ_4: $\Psi_1 - \Psi_3$	-.13	.8086	-1.7149
Ψ_5: $\Psi_2 - \Psi_3$	-.65	.9087	-2.4309
Subscale D			
Ψ_1: Black Train. vs. Control	.98	.5006	-.0962
Ψ_2: Hispanic Train. vs. Control	1.12	.6580	-.2946
Ψ_3: Anglo Train. vs. Control	1.00	.6658	-.4314
Ψ_4: $\Psi_1 - \Psi_3$	-.02	.8329	-1.6526
Ψ_5: $\Psi_2 - \Psi_3$	.12	.9360	-1.7146
Subscale E			
Ψ_1: Black Train. vs. Control	1.42	.4040	.5515*
Ψ_2: Hispanic Train. vs. Control	1.51	.5310	.3684*
Ψ_3: Anglo Train. vs. Control	1.31	.5373	.1549*
Ψ_4: $\Psi_1 - \Psi_3$	.11	.6722	-1.2075
Ψ_5: $\Psi_2 - \Psi_3$	.20	.7554	-1.2805

[a] See footnotes Table 11.

* Statistically significant at $\alpha_{\hat{\Psi}} = .033$

ing group scored significantly higher than their control group on subscales C and E. None of the interaction contrasts were significant.

The important thing to note is the pattern of treatment effects. It was hypothesized that training would be effective on the more difficult test items. In the case of the Raven subscales, which increase in average difficulty, and in the types of items comprising each subscale, it was expected that the latter subscales would be affected. A significant effect occurred for all groups on the most difficult subscale (E). For Hispanics, this was the only subscale on which a significant effect occurred. Of interest too, is that no effect occurred on subscale D for any of the groups.

Effect of Training on Raven Items Grouped According to M-demand

In order to examine the effects of training according to the item's complexity (i.e., information processing demands), the items were grouped according to M-demands as identified by Bereiter and Scardamelia (1979). According to Bereiter and Scaradamelia, Raven test items can be grouped according to their processing requirements. Table 13 gives the M-demand analysis which is taken from their Table 4 (Bereiter & Scardamelia, 1979, p. 60).

The means and standard deviations of the theoretically M-constructed subscales are presented in Table 14.

The analyses for each of the theoretically constructed Raven subscales are shown in Table 15.

The training effect was significant for Black students on all subscales except the most complex subscale, with an information processing demand (M-demand) of six. Hispanic students showed a significant training effect on items grouped with an M-demand of five. Anglo students showed

Table 13

Grouping of Raven Items According to M-Demand[a]

	M-Demand					
	1	2	3	4	5	6
RAVEN ITEMS	1	7	19	11	35	57
	2	8	21	12	36	58
	3	9	22	20	47	59
	4	10	23	24	48	60
	5	16	28	32	51	
	6	17	30	34	52	
	13	18	31	45	53	
	14	26	33	46	54	
	15	27	42	49	55	
	25	29	43	50	56	
	37	38	44			
		39				
		40				
		41				

---

[a] Numbers after underline and all 6-M items are based on p-values obtained on Anglo Control subjects participating in the present study.

Table 14

Average Score[a] and Standard Deviation of Raven Total and of Items Grouped According to M-Demand of Subjects with M-Level of Zero or Greater

M-Demand		Black		Hispanic		Anglo	
		Training	Control	Training	Control	Training	Control
ONE	$\bar{X}$	.96	.91	.97	.94	.99	.98
	SD	.08	.15	.05	.11	.05	.06
TWO	$\bar{X}$	.76	.65	.77	.70	.89	.87
	SD	.24	.27	.25	.27	.12	.20
THREE	$\bar{X}$	.67	.50	.63	.53	.79	.71
	SD	.29	.30	.31	.30	.17	.23
FOUR	$\bar{X}$	.48	.31	.49	.39	.62	.53
	SD	.27	.24	.30	.26	.17	.26
FIVE	$\bar{X}$	.19	.12	.27	.15	.35	.19
	SD	.20	.13	.19	.18	.24	.14
SIX	$\bar{X}$	.08	.06	.07	.05	.15	.05
	SD	.14	.11	.16	.12	.20	.10

[a] Percent passing

Table 15

Planned Contrasts for Selected Group Comparisons on Raven Items Grouped According to M-Demand

Comparison	$\hat{\Psi}$	$SE_{\hat{\Psi}}$	One-tail Confidence Interval[a] Lower Limit
M-Demand = 1			
Ψ_1: Black Train. vs. Control	.0486	.0177	.0105*
Ψ_2: Hispanic Train. vs. Control	.0338	.0233	-.0163
Ψ_3: Anglo Train. vs. Control	.0069	.0236	-.0438
Ψ_4: $\Psi_1 - \Psi_3$	.0417	.0295	-.0162
Ψ_5: $\Psi_2 - \Psi_3$	.0269	.0332	-.0063
M-Demand = 2			
Ψ_1: Black Train. vs. Control	.1091	.0433	.0160*
Ψ_2: Hispanic Train. vs. Control	.0704	.0569	-.0519
Ψ_3: Anglo Train. vs. Control	.0195	.0576	-.1043
Ψ_4: $\Psi_1 - \Psi_3$	.0896	.0720	-.0516
Ψ_5: $\Psi_2 - \Psi_3$	.0509	.0810	-.1078
M-Demand = 3			
Ψ_1: Black Train. vs. Control	.1670	.0505	.0584*
Ψ_2: Hispanic Train. vs. Control	.1032	.0664	-.0396
Ψ_3: Anglo Train. vs. Control	.0778	.0671	-.0665
Ψ_4: $\Psi_1 - \Psi_3$	.0892	.0840	-.0755
Ψ_5: $\Psi_2 - \Psi_3$	.0254	.0944	-.1597
M-Demand = 4			
Ψ_1: Black Train. vs. Control	.1696	.0461	.0705
Ψ_2: Hispanic Train. vs. Control	.1067	.0606	-.0235
Ψ_3: Anglo Train. vs. Control	.0936	.0613	-.0382
Ψ_4: $\Psi_1 - \Psi_3$	.0760	.0767	-.0744
Ψ_5: $\Psi_2 - \Psi_3$	.0131	.0862	-.1559
M-Demand = 5			
Ψ_1: Black Train. vs. Control	.0756	.0327	.0053*
Ψ_2: Hispanic Train. vs. Control	.1179	.0429	.0257*
Ψ_3: Anglo Train. vs. Control	.1627	.0434	.0694*
Ψ_4: $\Psi_1 - \Psi_3$	-.0871	.0543	-.1935
Ψ_5: $\Psi_2 - \Psi_3$	-.0448	.0611	-.1645
M-Demand = 6			
Ψ_1: Black Train. vs. Control	.0220	.0253	-.0324
Ψ_2: Hispanic Train. vs. Control	.0235	.0333	-.0481
Ψ_3: Anglo Train. vs. Control	.0925	.0337	.0200*
Ψ_4: $\Psi_1 - \Psi_3$	-.0705	.0421	-.1530
Ψ_5: $\Psi_2 - \Psi_3$	-.0690	.0473	-.1617

[a] See footnotes Table 11.

* Significant at α = .033

significant training effects on items with an M-demand of five and six. Again, none of the interactions was statistically significant.

The results indicate that training was most effective for Black students. The fact that Raven items with a M-demand of six were not affected for Blacks and Hispanics is not surprising since there were no subjects in either training group with an M-level of six.

The fact that none of the interactions was significant indicates that in those cases where training resulted in higher performance, the result was not greater for the minority students than for the Anglo students.

In general, the hypothesis that training would result in significant improvement was supported for the Black and Hispanic students on the total Raven. On Raven items grouped according to M-demand, only Black students showed consistent improvement. The hypothesis that training would lead to improved performance on the Raven did not hold for Anglo students.

The hypothesis that the gains for minorities would be significantly greater than for Anglo students was not supported at all. Although minority students showed significant differences over their control groups while the Anglo students did not, (which would lend some support to the statement that training was more effective for minority subjects). Overall this simply was not verified on the basis of the statistical interpretation of the statement applied in the analysis.

In summary, training was most effective for Black students on all of the analyses reported. Training was effective for Hispanics on the most difficult subscale of the Raven as well as total score.

Anglo training/control group differences were not significant on the Raven total, and with this one exception generally followed the pattern reported for Hispanics. Again, although the hypothesis of the effectiveness of training was supported for minority subjects, the hypothesis of greater effect for minority students was not supported. On the other hand, training was often effective for minority subjects, specifically Blacks, while it was not for Anglo students.

Finally, if one considers that the average M-level was generally less for minority subjects, and that the highest M-levels for minority control subjects was five, then training was effective on the items that are within their so-called processing capacity. The effect of training according to processing level is reported in the next section.

Effects of Training According to Processing Capacity

In this section the results of the effects of training according to subjects' M-level is reported. The results are reported for the Raven total and by theoretically constructed subscales (i.e., items grouped according to processing demands). The data are reported in two ways. First, Training effect results by M-level is reported for all subjects. This is then contrasted with Training effects when M-level is statistically controlled through analysis of covariance.

Raven Total by M-level

Results of Raven total test performance by M-level is given in Table 16. The first thing to note is the apparent monotonic relationship between average scores and increases in M-level. Perhaps more interesting, however, are the differences between groups within a given M-level. While there are some differences in total score (given small N's), the differences are not so great as to conclude that there are major differences overall between minority and Anglo subjects. That is, the main differences in the total scores across groups are due to differences in M-level within groups. This is interesting because M-level is a developmental variable,

Table 16

Average Raven Score of Control and Training Subjects by M-Level

M-Level		CONTROL			TRAINING		
		Black	Hispanic	Anglo	Black	Hispanic	Anglo
ZERO	$\bar{X}$	26.7	38.8	24.0	25.0	29.7	49.0
	SD	9.24	0	0	11.83	10.21	0
	N	13	1	1	6	3	1
ONE	$\bar{X}$	28.0	22.6	30.0	32.5	30.5	43.5
	SD	7.81	9.71	12.68	8.88	7.78	3.54
	N	7	5	4	13	2	2
TWO	$\bar{X}$	23.9	23.7	38.0	33.0	30.8	34.0
	SD	11.9	8.28	0	9.95	4.65	4.58
	N	9	7	1	6	4	3
THREE	$\bar{X}$	26.7	31.7	34.4	39.7	28.5	41.3
	SD	8.61	8.02	7.81	8.06	15.76	1.98
	N	13	6	7	12	6	7
FOUR	$\bar{X}$	31.5	31.8	40.9	42.9	39.1	40.2
	SD	8.94	10.31	5.32	6.64	4.67	4.62
	N	8	12	11	7	7	6
FIVE	$\bar{X}$	37.5	36.0	42.0	46.8	47.7	46.7
	SD	8.04	9.38	5.15	2.63	4.46	5.2
	N	12	9	12	4	7	9
SIX	$\bar{X}$	41.0	40.8	42.5	-	-	41.3
	SD	0	10.63	6.36	-	-	7.51
	N	1	4	2	-	-	3
TOTAL	$\bar{X}$	28.9	31.3	38.5	35.6	36.3	42.32
	SD	10.39	10.46	7.93	10.34	11.15	5.71
	N	63	44	38	58	29	31

vis-a-vis Piaget, which presumably increases with age. Since more minority subjects show lower M-levels, it is not surprising that they score lower on the Raven test. In this case one could conclude that the observed differences on the Raven are developmental, e.g., that minority subjects exhibit a lower, or "retarded" developmental rate (e.g., Jensen, 1974b).

On the other hand, the results for the training groups, which also show more minority subjects with lower M-levels, suggests that there is little within M-level difference in Raven performance after training. This is particularly true at M-level greater than one, and more so for Black subjects. Thus, while group differences on the Raven test are related to differences in M-level, it is not because of retarded development. A purely "developmental lag" explanation does not explain these results since cognitive developmental growth generally takes longer than two weeks, and is usually unaffected by training. Since there is little difference in Raven scores within M-levels for Training subjects it is more likely that minority subjects simply do not comprehend the task requirements of the Raven test in the first place.

Group Comparison with Anglo Control Students

Analysis of the data in Table 16 follows from hypothesis that training would be more effective for minority subjects. Stated another way, the usual observed differences between minority and Anglo students would be reduced or eliminated by training.

The ANCOVA was computed through regression analysis by dummy coding group membership and treating the Anglo control group as the references category. In this way comparisons between each group and the Anglo control are output directly from the analysis. With this procedure there are five comparisons, each tested at alpha equal .02 for each contrast and .10 over all comparisons.

An alpha of .10 was selected since the complete factorial design (ethnic group by treatment) allows for an alpha of .15. However, since an assumption of ANCOVA is homogeneity of regression and because the hypothesis does not technically require a group-by-treatment analysis, the remaining .05 was allocated to a statistical test of the homogeneity assumption. The test for homogeneity of regression is reported first, followed by the results of the ANCOVA.

Test for Homogeneity of Regression. The statistical test for parallelism of regression lines was performed according to the procedures described in Kerlinger and Pedhauzer (1973). In their procedure, an F-test is performed on the difference in R^2 obtained for the regression of Raven onto group membership and separate vectors representing M-level for each group, and the R^2 for Raven regressed onto group membership and a single vector representing M-level. The F formula is:

$$F = \frac{(R^2{}_1 - R^2{}_2)/k_1 - k_2}{(1 - R^2 1)/N - k_2 - k_2}$$

Where: $R^2{}_1$ = Raven onto group and M-level vectors for each group.
$R^2{}_2$ = Raven onto group and M-level.
k_1 = Number of vectors for $R^2{}_1 = 11$
k_2 = Number of vectors for $R^2{}_2 = 6$

The computed R^2's are .384 for Raven onto group and separate M-level vector and .375 for Raven onto group and a single M-level vector. For these values F = .687, which is not significant for $F_5 = 2.21$. Thus, the assumption of parallelism is supported.

ANCOVA Results. The results for the ANCOVA controlling for M-level across all groups are summarized in Table 17. The results show that when each group is compared with the Anglo control group (as criterion) significant differences occur between minority Control and Anglo Control group students whereas no differences occur for minority Training group students and Anglo Control group students, nor do differences occur betwen the Anglo Training and Control groups. Clearly, training had the effect of eliminating the initial differences between minority and Anglo students.

Group Comparisons on Theoretically Defined Raven Subscales

Homogeneity of Regression. The same analysis was performed for each theoretically constructed Raven subscale. The statistical test for parallelism for each subscale is reported in Table 18.

The R^2's in Table 18 represent regressions of each subscale onto group membership (defined by the six groups described above) and a single vector representing M-level, and regressions of the subscales onto group membership and separate vectors representing M-level according to group membership. The difference in R^2's is tested by an F-test and represents the test for homogeneity of regressions among the groups (Kerlinger and Pedhauzer, 1973).

Results of the F-test necessitates rejection of the hypothesis of parallel regression lines for the "Three-M" subscale and indicates that analysis of covariance is inappropriate. The remaining five subscales are not significant, so an ANCOVA was performed for these subscales. Table 19 presents the ANCOVA results.

ANCOVA. The data in Table 19 represent significant tests between five of the groups and the Anglo Control group. Beta (B) is the deviation of each group (i.e., effect) from the mean score obtained on the given sub-

Table 17

Summary ANCOVA for Comparisons of Study Groups with Anglo Control Group Subjects on Raven Test with M-Level as Covariate

Variable	B	SE	F
M-Level	3.0141	.3314	82.74*
Black Training	1.7512	1.8389	.91
Hispanic Training	-0.3134	2.0977	.02
Anglo Training	3.9128	2.0489	3.65
Black Control	-5.9611	1.7831	11.18**
Hispanic Control	-6.4305	1.8768	11.74**
Anglo Control (Constant)	27.1313		

* Significant at $p < .05$, $F \geq 2.21$
** Significant at $p < .02$, $F \geq 5.38$

Table 18

Homogeneity Test for the Regression of Raven M-Demand Subscales onto M-Level in Black, Hispanic and Anglo Training and Control Groups

Raven Subscale	R^2 (M-Level/Group)	R^2 (Group/[Group x M-Level])	F^a
M-Demand=1	.13524	.15200	0.99
M-Demand=2	.26229	.28767	1.79
M-Demand=3	.24867	.28811	2.78*
M-Demand=4	.36711	.37484	0.62
M-Demand=5	.26594	.28765	1.58
M-Demand=6	.07766	.10280	1.21

[a]

$$F = \frac{(R^2 - R^2)/(11 - 6)}{(1 - R^2)/263 - 11 - 1}$$

(Kerlinger and Pedhauzer, 1974).

* Significant $p < .05$, $F_{5,\infty} = 2.21$

Table 19

Summary ANCOVA of Study Groups with
Anglo Control Group on Raven M-Demand Subscales: M-Level as Covariate

Subscale M-Demand[a]	Variable	B	SE	F
M=1	M-Level	.1714	.0407	17.78*
	Black Training	.0351	.2257	0.02
	Hispanic Training	.0336	.2573	0.02
	Anglo Training	.0792	.2513	0.10
	Black Control	-.5589	.2187	6.53**
	Hispanic Control	-.3999	.2302	3.02
M=2	M-Level	.8463	.1194	50.264*
	Black Training	-.2701	.6625	0.17
	Hispanic Training	-.9228	.7557	1.49
	Anglo Training	.2913	.7381	0.15
	Black Control	-2.0909	.6424	10.59**
	Hispanic Control	-2.2151	.6761	10.73**
M=4	M-Level	.8161	.0853	91.52*
	Black Training	.7424	.4735	2.46
	Hispanic Training	.1518	.5401	0.08
	Anglo Training	.9536	.5275	3.27
	Black Control	-1.2375	.4591	7.27**
	Hispanic Control	-1.2111	.4832	6.28**
M=5	M-Level	.4219	.0653	41.79*
	Black Training	.6593	.3622	3.31
	Hispanic Training	1.0667	.4132	6.67**
	Anglo Training	1.6363	.4036	16.44**
	Black Control	-0.2437	.3512	0.48
	Hispanic Control	-0.2647	.3697	0.51
M=6	M-Level	.0650	.0214	9.19*
	Black Training	.1998	.1189	2.82
	Hispanic Training	.1059	.1357	0.61
	Anglo Training	.3715	.1325	7.86**
	Black Control	.0891	.1153	0.60
	Hispanic Control	-.0116	.1214	0.01

[a] The M-Demand = 3 subscale did not meet the homogeneity assumption so ANCOVA was not performed for this subscale.

* Significant at $p < .05$, $F > 2.21$

** Significant at $p < .02$, $F > 5.38$

scale by the Anglo Control group subject. The F-value is tested for significance at alpha equal .05 for the covariant (M-level) and .02 for each of the group effects.

The covariant, M-level, was significant for all subscales. Black control group students scored significantly different than Anglo control students on subscales with M-demands of one, two and four, but not on the Five-M and Six-M demand subscales. Black training group subjects showed a significant difference on the Five-M subscale only. Anglo training group students scored significantly different on the Five-M and Six-M subscales.

The direction of the differences was such that minority control groups always scored lower than Anglo control students while training students in all ethnic groups scored higher. These results indicate that when M-level is controlled, differences between minority and Anglo control stu- dents-occurs primarily on the less complex subscales. No differences occur on the most complex subscales except in one instance where Hispanic students scored higher. In all other instances, training group subjects scored higher, although the differences are not significant. This finding is in contrast to reports that minority subjects perform lower on more complex tasks when compared to Anglo students (Jensen, 1974a, 1980). When developmental level is controlled, there are no differences in performance. Moreover, given training, it is apparent by observation of mean scores in Table 16 that there are no differences between minority and Anglo students.

Overall, the results indicate that the main difference in performance on the Raven is due to differences in performance on the developmental measure. This might lead one to conclude that the difference is "developmental lag." However, observation of the mean scores within a given M-level together with the effect of training suggest that this is not the case. The most appealing argument is that there is a general "test taking skill" reflected both in the M-level measure and in the performance of Control group subjects. Training provided the test taking experience necessary to produce group parity in performance.

CHAPTER IV

RESULTS: CULTURE-LOADING

Examination of the Culture-Loaded Hypothesis

Results of the culture-loading hypothesis are presented in three sections. In the first section the procedure for identifying culture-loaded items and its assumptions are described. The results of applying the procedure are then presented. Since different items were identified as culture-loaded, results are reported separately for Black and Hispanics. The second section examines the validity of the procedure by comparing group performance against outcomes considered to be consistent with a culture-loaded interpretation. In this section expected outcomes are first described, then an anlysis of group performance is presented at the individual item level and for items grouped according to processsing demands. Results are reported separately for Blacks and Hispanics. The final section is a discussion of the results of the study.

Identification of Culture-Loaded Test Items

Procedure

Items are identified as culture-loaded when there are differences in the information processing demands of the item for minority control subjects in comparison to Anglo control group subjects. The criteria for determining whether the information processing demand of an item is different for minority subjects is made by examining the percentage passing at the M-level equal to the M-demand of the item. If there is a 25% or more difference between minority and Anglo subjects of the same M-level, then the item is determined to require a greater processing capacity for minority subjects and hence is identified as culture-loaded.

In a few instances minority processing demands were nearly equal to that of Anglo subjects, but differences equal to or greater than 25% occurred at higher M-levels. In these cases items were also judged to require a greater processing demand, i.e., were culture-loaded. This was done separately for Black and Hispanic students.

The procedure is consistent with the theoretical discussion provided in previous chapters. That is, that culture-loading effects the processing demands of a task, and that differences in processing requirements can occur because of difference in processsing strategies, or because of over sensitivity to misleading cues. The difference, however, is hypothesized to be due to experience rather than to ability per se. In either case, the processing requirements of the task are affected.

Total percent passing for each group is ignored in this process in favor of group differences within M-levels. Thus, culture-loaded items may or may not show ethnic group differences. Total percent passing is not important for two reasons. First, it assumes a priori that no group differences should exist. Second, differences are expected since the number of subjects at each M-level is generally not the same across groups. That is, group differences in percent passing are not meaningful from a developmental perspective since the groups were not matched on developmental variables.

In contrast, the assumption of no group differences within M-levels in order to identify items as culture-loaded is based on the theoretical position that subjects of the same M-level are developmentally equal. Thus, the argument is that one can assume that no group differences should exist. The reasonableness of this assumption, of course, rests with the outcome of the analysis reported in this part of the study.

A further assumption in the procedure for identifying culture-loaded test items is that the FIT itself is unbiased. While this assumption may be questioned, it is clear that the finding of group differences alone is not sufficient. Moreover, if the FIT is biased, the most likely outcome would be that minority subjects' M-level is really higher than indicated. This means that Anglo subjects are being compared with minority subjects who can process more information. However, this would simply lead to less items being identified as culture-loaded and would serve to provide a more conservative test of the hypothesis.

Finally, since there are no, or very few minority training group subjects with an M-level of six, items with an M-demand of six could not be included as part of the examination of culture-loading.

Research Questions

The research issues examined in this chapter are:

1. Raven test items identified, a-priori, as culture-loaded will show greater effect due to training than items not identified as culture-loaded. That is, training will have its greatest effect on the items identified as culture-loaded.
2. Greater group differences will be found on culture-loaded than on non-culture loaded test items.

3. The training effect for minority students on culture-loaded items will be greater than the training effect for Anglo students.
4. Training will reduce or eliminate group differences on culture loaded test items; the effect will be greater than for non-culture loaded test items.

Results

Application of the criteria described above resulted in 26 items identified as culture-loaded for Blacks and Hispanics. In most instances the same items were identified as culture-loaded. There were, however, a few deviations from this pattern. Culture-loaded items are shown, according to M-demand, in Table 20. The letter next to the item indicates the original Raven subscale to which the item belongs.

The total number and proportion of items identified as culture-loaded within each M-demand subscale for Black and Hispanic students are also shown in Table 20. The highest percentage of culture-loaded items were found in subscales that are within the processing capacity (M-level) of the students. Items with M-demands of one showed few items as culture-loaded and were probably influenced by a ceiling-effect in that the majority of subjects in each ethnic group were able to solve the items.

Examination of the items according to original Raven subscales indicates that items of average to above average item difficulty were more often identified as culture-loaded. This observation, together with the item's M-demand suggest that floor and ceiling-effects are a factor in identifying culture-loaded items according to the criteria used. That is, items within the processing and difficulty level of the subjects are more likely to be identified as culture-loaded. This is not to say that very

Table 20

Culture-Loaded Raven Test Items for Black and Hispanic Students by M-Demand

		M-Demand[a]					
		1	2	3	4	5	
BLACK		6 (A) 14 (B) 15 (B)	7 (A) 9 (A) 10 (A) 16 (B) 17 (B) 18 (B) 26 (C) 29 (C) 38 (D) 39 (D) 41 (D)	19 (B) 21 (B) 28 (C) 31 (C) 43 (D)	11 (A) 12 (A) 24 (B) 32 (C) 49 (E) 50 (E)	54 (E)	
	N	3	11	5	6	1	26
	%	27	79	45	60	10	46
HISPANIC		15 (B)	7 (A) 8 (A) 9 (A) 10 (A) 16 (B) 17 (B) 18 (B) 29 (C) 38 (D) 39 (D) 40 (D)	19 (B) 23 (B) 30 (C) 31 (C) 42 (D) 43 (D)	12 (A) 24 (B) 32 (C) 45 (D) 46 (D) 50 (D)	52 (E) 54 (E)	
	N	1	11	6	6	2	26
	%	09	79	55	60	20	46

[a] Since there were no minority training group subjects with an M-Level of six, items with an M-Demand of six could not be included.

easy or very difficult items are not culture-loaded. Rather, they simply do not provide a range in which the particular criteria had enough power to detect.

In all, this means that there are probably some items that were not identified as culture-loaded. For this reason and because of possible bias in the FIT, it is probable that not all culture-loaded items were identified.

Validation of the Culture-Loading Hypothesis

Outcomes Consistent with a Culture-Loaded Interpretation

Basically a culture-loaded hypothesis is one in which test performance is in part a function of characteristics unique to the group but not necessarily related to what the test is supposed to measure in the first place. Once items are identified, outcomes can be specified which, if verified, would support a culture-loaded interpretation and hence a culture-loaded hypothesis. There are three basic research hypotheses of interest in the evaluation of the validity of identifying culture-loaded items on the basis of the proposed theoretical information processing model. Positive findings would support the approach used in this study and a culture-loaded interpretation.

One expected outcome is that greater observed group differences would occur on items identified as culture-loaded than on non-culture loaded items. A second expectation is that training would effect culture-loaded items more than non-culture loaded items. Finally, it is expected that training would be more effective for minority students than for Anglo students on the culture-loaded items but not necessarily on the non-culture loaded items.

Examination of the expected outcomes is made by comparing various differences in performance between minority and Anglo groups, and Training and Control groups on culture-loaded and non-culture loaded items individually, and on the subscale totals when the items are grouped according to M-demand. Five group comparisons are related to the expected outcomes: 1) between minority and Anglo control groups; 2) between minority and Anglo training groups; 3) between training and control subjects in the minority group; 4) between training and control subjects in the Anglo group; and, 5) the interaction of treatment by ethnicity.

The comparisons correspond to specific culture-loading expectations. Comparisons one and two concern expectations about the effect of training on group differences. It is expected that culture-loaded items would show greater between ethnic group differences than non-culture loaded items. Items which show such a trend would be consistent with the culture-loading expectation.

Comparisons three and four concern the expectation that training would have a greater effect on minority group differences (i.e., between training and control groups) than on Anglo group differences. Thus, it would be expected that training would be more effective for minorities on culture-loaded items than on non-culture loaded items. The same outcome would not necessarily be expected for Anglo comparisons.

Finally, comparison five concerns the expectation that training would be more effective for minorities than for Anglo students on the culture-loaded items. The only statement made concerning the expected effect of training for Anglo students on non-culture loaded items, is that it should be more or less equal to minority subjects. This will be discussed in more detail once the data has been examined. Table 21 provides a summary of the

Table 21

Summary of Culture-Loaded Outcomes
for Specific Group Comparisons

Comparison	Culture Loade Items	Non-Culture Loaded Items
ONE Control: Minority-Anglo	Should be relatively large and less than Zero (i.e., negative)	May or may not be large and negative since items may show group differences and not be culture loaded.
TWO Training: Minority-Anglo	Should be smaller than comparison One.	Should be the same or less than comparison One or non-culture loaded items and greater than comparison Two for culture loaded items.
THREE Minority: Train-Control	Should be large and positive, and greater than corresponding comparison for non-culture loaded items.	Should be less than corresponding comparison for culture loaded items. If not smaller than comparison Four should be greater than Zero.
FOUR Anglo: Train-Control	Should be smaller than corresponding comparison for minority students.	No prediction
FIVE Interaction	Should be relatively large (greater than 4%) and positive. Will be greater than comparison Five for non-culture loaded items.	Relatively small and less than comparison Five for culture loaded items.

outcomes expected on the five comparisons which would be consistent with a culture-loaded hypotheses.

In order to test these outcomes, percent passing each item for each ethnic group was computed for subjects whose M-level was equal or greater to the M-demands of the items. The rationale for this is that comparisons between groups, and the effects of training, would be examined only on those subjects who possess the minimum processing resources required for a given item. This takes into account, somewhat, the discrepancy in proportion of minority subjects obtaining lower M-levels on the FIT. More important, however, is that it reflects the position that training would most likely be effective for those subjects who possess the minimum resources in the first place.

Examination of the expected outcomes is made both for culture-loaded and non-culture loaded items individually, and for culture-loaded and non-culture loaded subscale totals formed by grouping the items according to M-demand. Because of the number of statistical tests required, examination of individual items is made by observation of the pattern of differences in performances. Analysis of variance with planned comparisons is used to examine the hypothesis of interest for the total score obtained when culture-loaded and non-culture loaded items are grouped according to M-demand. In the following the results of individual item data are presented first. Following this, analyses of culture-loaded and non-culture loaded subscales are statistically examined.

Item Analysis for Culture-Loading

Individual item results are examined by M-demand for Blacks and Hispanics separately. Item data is reported according to the previously defined five group comparisons in Tables 22 and 23 for Blacks and Hispanics

Table 22

Comparison on Culture-Loaded and Non-Culture Loaded Raven Items for Black and Anglo Training and Control Groups

Culture Loaded

Items	Comparison[a]				
	1	2	3	4	5
M-Demand=1					
6	-.10	.09	.22	.03	.10
14	-.11	.13	.23	.00	.11
15	-.17	.01	.10	-.05	.15
M-Demand=2					
7	-.21	.00	.10	-.11	.21
9	-.19	-.10	.13	.03	.09
10	-.25	-.21	.05	.01	.04
16	-.24	.18	.33	-.10	.42
17	-.23	-.13	.03	-.08	.10
18	-.37	.04	.15	-.26	.41
26	-.28	-.35	-.08	.00	-.08
29	-.25	.19	.08	.01	.07
38	-.15	-.20	.05	.10	-.05
39	-.23	-.16	.13	.05	.07
41	-.13	-.15	.08	.10	-.02
M-Demand=3					
19	-.22	-.07	.12	-.04	.15
21	-.09	-.07	.16	.14	.02
28	-.16	.15	.24	-.07	.31
31	-.35	-.13	.48	.27	.22
43	-.18	-.10	.05	.03	.08
M-Demand=4					
11	-.21	-.20	.06	.05	.01
12	-.38	.11	.43	-.05	.49
24	-.26	-.13	.32	.19	.13
32	-.21	.05	.31	.05	.26
49	-.47	.02	.35	-.14	.49
50	-.19	-.11	.33	.25	.08
M-Demand=5					
54	-.25	.44	.83	.14	.69

Non-Culture Loaded

Items	Comparison				
	1	2	3	4	5
M-Demand=1					
1	.00	.04	.04	.00	.04
2	.00	.04	.04	.00	.04
3	-.01	.00	.04	.03	.01
4	.03	.00	.00	.03	-.03
5	-.05	-.02	.09	.06	.03
13	-.02	.04	.06	.00	.06
25	-.09	-.11	.04	.06	-.02
37	-.06	-.07	-.02	-.01	-.01
M-Demand=2					
8	-.14	-.10	.08	.03	.04
27	-.06	-.13	.10	.16	-.06
40	-.16	-.12	.03	-.07	.04
M-Demand=3					
22	.01	.01	.13	.12	.00
23	.01	.02	-.04	-.05	.02
30	-.09	-.07	.13	.11	.02
33	-.09	.01	.33	.24	.09
42	-.16	-.17	.11	.12	-.01
44	.05	.07	.13	.10	.03
M-Demand=4					
20	-.01	.06	.23	.16	.07
34	.04	-.45	.02	.48	-.46
45	.21	.14	-.18	-.12	-.06
46	-.11	-.33	.04	.26	-.23
M-Demand=5					
35	.17	-.53	-.17	.53	-.69
36	.00	.03	.17	.14	.03
47	.17	-.22	-.33	.06	-.39
48	.00	.14	.25	.11	.14
51	.00	-.17	.25	.42	-.17
52	.00	-.03	.33	.36	-.03
53	-.08	.11	.75	.56	.19
55	-.08	-.22	-.08	.06	-.14
56	-.17	-.31	.17	.31	-.14

[a] 1 = Control: Black-Anglo
2 = Training: Black-Anglo
3 = Black: Training-Control
4 = Anglo: Training-Control
5 = Interaction: Treatment by Ethnicity

Table 23

Comparison on Culture-Loaded and Non-Culture Loaded Raven Items for Hispanic and Anglo Training and Control Groups

Culture Loaded

Items	Comparison[a] 1	2	3	4	5
M-Demand=1					
15	-.25	.00	.21	-.05	.26
M-Demand=2					
7	-.26	-.08	.07	-.11	.18
8	-.20	-.25	-.01	.03	-.05
9	-.09	.00	.12	.03	.09
10	-.08	.04	.05	.01	.04
16	-.23	-.12	.01	-.10	.11
17	-.29	.01	.22	-.08	.29
18	-.31	.06	.11	-.26	.37
29	-.29	-.12	.17	.01	.16
36	-.29	-.21	.17	.10	.08
39	-.37	-.16	.26	.05	.21
40	-.35	-.24	.04	-.07	.10
M-Demand=3					
19	-.32	-.31	-.03	-.04	.01
23	-.21	-.01	.14	-.04	.20
30	-.26	-.21	.16	.11	.05
31	-.14	-.13	.26	.27	-.01
42	-.24	-.25	.11	.12	-.01
43	-.18	.18	.33	-.03	.36
M-Demand=4					
12	-.31	.00	.26	-.05	.31
24	-.28	-.21	.26	.19	.07
32	-.22	-.14	.12	.05	.07
45	-.14	-.07	-.05	-.12	.07
46	-.50	-.14	.62	.26	.36
50	-.18	.07	.50	.25	.25
M-Demand=5					
52	-.19	-.06	.49	.36	.13
54	-.31	.16	.60	.14	.46

Non-Culture Loaded

Items	Comparison 1	2	3	4	5
M-Demand=1					
1	.00	.00	.00	.00	.00
2	.00	.00	.00	.00	.00
3	.03	-.04	-.04	.03	-.07
4	.00	-.04	-.01	.03	-.04
5	.03	.00	.03	.06	-.03
6	-.02	.00	.05	.03	.02
13	-.03	.00	.03	.00	-.03
14	-.05	.00	.05	.00	.05
25	-.12	-.12	.06	.06	.01
37	-.07	.00	.06	-.01	.08
M-Demand=2					
26	-.18	-.29	-.12	.00	-.12
27	-.07	-.29	-.06	.16	-.22
41	-.14	-.17	.07	.10	-.03
M-Demand=3					
21	-.04	-.06	.12	.14	-.02
22	-.03	-.06	.10	.12	-.03
28	-.13	-.11	-.05	-.07	.02
33	-.07	-.05	.26	.24	.02
44	-.01	-.07	.04	.10	-.06
M-Demand=4					
11	.12	.07	0	.05	-.05
20	-.12	-.07	.21	.16	.05
34	-.11	-.36	.24	.48	-.24
49	-.07	.07	.00	-.14	.14
M-Demand=5					
35	-.14	-.49	.17	.53	-.35
36	.03	.08	.03	.14	-.11
47	-.06	-.22	-.11	.06	-.17
48	.00	-.11	.00	.11	-.11
51	.19	.33	.56	.42	.14
53	.00	-.17	.38	.56	-.17
55	-.06	.21	.32	.06	.26
56	-.03	.16	.49	.31	.19

[a] 1 = Control: Hispanic-Anglo
2 = Training: Hispanic-Anglo
3 = Hispanic: Training-Control
4 = Anglo: Training-Control
5 = Interaction: Treatment by Ethnicity

respectively. The comparisons in Tables 22 and 23 represent differences in percent passing when a subjects' M-level in each group are matched with the M-demand of the item.

Since statistical analyses are not usually based on item data, a criteria of 4% was arbitrarily selected as demonstrating a difference in any particular comparison. Thus, if a treatment effect is expected to be positive and relatively large, it had to be at least 4% to be considered as supportive. Similarly, if a comparison is anticipated to show no effect, it had to be less than 4%. In the following discussion this criteria is applied when judging whether a particular outcome is consistent with a culture-loaded expectation. Results are presented for Blacks first, then Hispanics.

Blacks: Observation of Table 21 indicates that all of the items with an M-demand of one followed the expected pattern. First, larger group differences are found on culture-loated items than on non-culture loaded items. Second, the group difference on culture-loaded items is reduced through training. This did not occur for non-culture loaded items. Third, the effect of training was greater on culture-loaded items than on non-culture loaded items. Finally, the gains for Black students are larger than for Anglo students.

For items with an M-demand of two, 10 of 14 follow the expected direction in support of the culture-loaded expectation. Nine of 11 of the Three-M-demand items show a pattern consistent with culture-loading and 8 of 10 of the Four-M and the Five-M items come out as predicted. Two apparent culture-loaded items were not detected on the Five-M subscale.

Overall, 46 (or 82%) of the 56 items examined showed results consistent with a culture-loaded expectation. Of the 10 items that did not conform to the model, four non-culture loaded items produced results which suggested they were misidentified. Three culture-loaded items produced results somewhat consistent with a culture-loaded classification, but yielded interaction gains in favor of minorities of less than 4%. Thus, these were not counted as supportive of the hypothesis. The remaining three items were culture-loaded but produced results opposite of expectation. Two showed interactions in favor of Anglo students (5% and 8%), and one showed no difference (2%) instead of a gain for minority students.

Hispanics: Results for Hispanics (Table 22) reveal that One-M items produce outcomes consistent with the culture-loaded interpretation. Two non-culture loaded items, however, show a 5% and 8% gain for minority students. All but two of the Two-M items appear correctly identified. The two errors were items identified as culture-loaded. One showed a 5% gain for Anglos and the other, while consistent with the culture-loaded expectations, showed only a 4% gain for Hispanic students over Anglos. Three of the 11 Three-M items did not follow the expected pattern. Three culture-loaded items showed no gain for minorities even though they were consistent on the other comparisons.

Of the 10 Four-M items, two non-culture loaded items showed gains of 5% and 14% in favor of Hispanics, inconsistent with their identification. Three non-culture loaded items in the Five-M subscale produced results consistent with a culture-loaded identification.

Overall, 44, or 79% of the 56 items produced patterns consistent with a culture-loaded hypothesis. Seven of the 12 misidentified items were classified as non-culture loaded when in fact they produced results expected of culture-loaded items. The gain for these items ranged from 5% to 20% in favor of Hispanics. The remaining four errors were identified as culture-loaded. Of these, two produced gains in favor of Anglos and two showed no gain for either group.

In general, the results support a culture-loaded explanation and thus support the definition based on information processing capacity. In most cases items identified as culture-loaded and non-culture loaded produced results in the desired directions. In several of the items that were misidentified, many either showed no gain for minority students or were thought to be non-culture loaded but produced a large gain in favor of minority students. These outcomes are consistent with the previous statement that the procedure for identifying culture-loaded items is probably conservative, i.e., is likely to miss some culture-loaded items.

Summary

The above discussion was based on subjective judgements about the expected size of the differences in comparisons. A conservative criteria was applied, although some may certainly wish to argue this point. The subjective rule was to judge items as agreeing with the culture-loaded hypothesis if: a) results on all 5 comparisons were in the predicted direction, b) the gains for minority students on culture-loaded items were greater than 4% over Anglo gains, and c) the gains for minority students on non-culture loaded items was 4% or less. If any of these three criteria were not satisfied, the results were not considered to support a culture-loaded interpretation. The main point, however, is that the focus was mainly on the constancy of the pattern produced. To this extent there was certainly a constancy observed. The overall results are summarized in Table 24.

In general the procedure for identifying culture-loaded items appeared to be consistent with expected outcomes. Items identified as culture-loaded: 1) produced greater group differences than non-culture loaded

Table 24

Summary Results of Agreement[a] in Identification of Culture Loaded and Non-Culture Loaded Raven Test Items

BLACKS			
	Met Expectation		
Predicted	Culture Loaded Items	Non-Culture Loaded Items	Total
Culture Loaded	20	6	26
Non-Culture Loaded	5	25	30
Total	25	31	56

HISPANICS			
	Met Expectation		
Predicted	Culture Loaded Items	Non-Culture Loade Items	Total
Culture Loaded	21	5	26
Non-Culture Loaded	7	23	30
Total	28	28	56

---

[a] Underline indicates agreement between items predicted as culture-loaded or non-culture loaded and those that met the expected criteria.

items; 2) had the difference reduced or eliminated through training; 3) showed greater training effects; and 4) produced greater gains for minority students due to training than for Anglo students. Non-culture loaded items generally produced the opposite results.

The following section presents the application of a statistical criteria to the results of culture-loaded and non-culture loaded subscales.

Analysis of Culture-Loading on Items Grouped According to M-demand

Statistical tests of the culture loading hypothesis were made by grouping items of the same M-demand into two subscales according to whether the items are culture-loaded or not culture-loaded. An analysis of the five comparisons described above was then performed. The Cell-means model of analysis was used to test the statistical significance of the five contrasts. The "mean squares within" term for each subscale was obtained through a "one-way" ANCOVA between the particular groups included in the analysis i.e., Black-Anglo and Hispanic-Anglo. Contrasts were computed separately because different items were identified as culture-loaded for Black and Hispanic groups.

Between group contrasts were tested with alpha controlled at .025 (one-tail) yielding a total alpha of .10. The interaction contrast was tested at alpha equal to .05 (one tail). This is consistent with a .15 alpha allowed in the full factorial two-way analysis of variance with interaction. In the following, the results are reported separately, first for Blacks then Hispanics.

Blacks. Table 25 gives the percent of Blacks and Anglos passing culture-loaded and non-culture loaded subscales defined according to M-demand.

Statistical analysis of the data is provided in Table 26. The data in Table 26 show the difference between percent passing for each comparison,

Table 25

Percent Passing Culture-Loaded and Non-Culture Loaded Items for Black and Anglo Training and Control Groups

M-Demand		Culture Loaded				Non-Culture Loaded			
		Control		Training		Control		Training	
		Black	Anglo	Black	Anglo	Black	Anglo	Black	Anglo
ONE	$\bar{X}$	.86	.98	1.00	.97	.95	.98	.98	.99
	SD	.28	.08	.00	.09	.11	.07	.08	.03
(1≤M-Level≤5)	N	46	54	54	25	46	35	54	25
TWO	$\bar{X}$	.69	.92	.79	.90	.79	.91	.84	.96
	SD	.28	.11	.24	.12	.26	.17	.25	.11
(2≤M-Level≤5)	N	40	31	40	23	40	31	40	23
THREE	$\bar{X}$	.54	.73	.75	.81	.74	.73	.84	.81
	SD	.29	.26	.25	.17	.38	.32	.27	.25
(3≤M-Level≤5)	N	31	30	23	21	31	30	23	21
FOUR	$\bar{X}$	.39	.68	.69	.74	.58	.55	.60	.75
	SD	.23	.24	.25	.17	.32	.30	.29	.17
(4≤M-Level≤5)	N	19	23	12	14	19	23	12	14
FIVE	$\bar{X}$	.17	.42	1.00	.56	.63	.56	.69	.78
	SD	.39	.51	0	.53	.35	.26	.24	.20
(M-Level = 5)	N	12	12	4	9	12	12	4	9
TOTAL	$\bar{X}$	.56	.77	.71	.80	.54	.60	.61	.72
	SD	.22	.18	.27	.11	.15	.13	.16	.10
	N	46	35	54	25	46	35	54	25

Table 26

Group Comparisons on Culture Loaded and Non-Culture Loaded Items for Black and Anglo Training and Control Group Students

	Culture Loaded				Non-Culture Loaded			
	$\hat{\psi}$	SE $\hat{\psi}$	τ	η^2	$\hat{\psi}$	SE $\hat{\psi}$	τ	η^2
M-Demand=1 (1≤M-Level≤5)								
Control: Black-Anglo	-.13	.0832	-1.51	-	-.03	.0182	-1.46	-
Training: Black-Anglo	.07	.0897	.779	-	-.01	.0196	-.57	-
Black: Training-Control	.19	.0744	2.53*	5.0	.04	.0163	2.18*	2.9
Anglo: Training-Control	-.01	.0971	-0.078		.02	.0213	0.94	-
Interaction	.20	.1223	1.60	2.0	.02	.0268	0.58	-
M-Demand=2 (2≤M-Level≤5)								
Control: Black-Anglo	-.23	.0519	-4.44**	12.9	-.12	.0523	-2.34**	3.9
Training: Black-Anglo	-.11	.0568	-2.02**	2.6	-.11	.0572	-2.01**	2.9
Black: Training-Control	.09	.0485	1.92	2.4	.05	.0489	1.02	-
Anglo: Training-Control	-.02	.0597	-0.38	-	.04	.0602	0.71	-
Interaction	.12	.0769	1.50	-	.01	.0775	0.10	-
M-Demand=3 (3≤M-Level≤5)								
Control: Black-Anglo	-.20	.0687	-2.89**	6.1	.04	.0721	- .240	-
Training: Black-Anglo	-.04	.0809	-.537	-	.02	.0849	- .617	-
Black: Training-Control	.21	.0738	2.82**	5.8	.13	.0774	1.69	2.0
Anglo: Training-Control	.05	.0763	.689	-	.11	.0800	1.33	-
Interaction	.15	.1062	1.46	-	.02	.1114	0.22	-
M-Demand=4 (4≤M-Level≤5)								
Control: Black-Anglo	-.29	.0706	-4.06**	18.5	.02	.0882	.28	-
Training: Black-Anglo	-.04	.0895	-0.49		-.15	.1120	-1.30	2.5
Black: Training-Control	.30	.0839	3.57*	14.3	.03	.1049	0.24	-
Anglo: Training-Control	.06	.0771	0.74		.20	.0965	2.03*	6.0
Interaction	.24	.1140	2.13***	5.1	-.17	.1425	-1.20	2.1
M-Demand=5 (M-Level=5)								
Control: Black-Anglo	-.25	.1854	-1.35	4.1	.00	.0801	0.00	-
Training: Black-Anglo	.44	.2729	1.63	6.0	-.13	.1178	-1.13	2.7
Black: Training-Control	.83	.2622	3.18*	23.0	.15	.1132	1.31	3.7
Anglo: Training-Control	.14	.2002	0.69	-	.28	.0865	3.25*	22.5
Interaction	.69	.3299	2.11***	10.0	-.13	.1425	-0.93	-
All Items (1≤M-Level≤5)								
Control: Black-Anglo	-.21	.0488	-4.30**	10.0	-.06	.0315	-1.9[illegible]	2.1
Training: Black-Anglo	-.09	.0526	-1.69		-.10	.0340	-3.11*	5.3
Black: Training-Control	.16	.0436	3.57*	6.9	.07	.0282	2.49*	3.4
Anglo: Training-Control	.03	.0570	0.61		.11	.0368	3.12*	5.4
Interaction	.12	.0718	1.69***	1.5	-.04	.0463	-0.96	-

* Significant at $p < .025$, $t \geq 1.96$

** Significant at $p < .025$, $t \leq 1.96$

*** Significant at $p < .05$, $t \geq 1.645$

standard error term, t-value, and the percent of variance explained by the contrast. The data are organized according to item M-demand and represent performance of subjects with an M-level equal to or greater than the M-demand of the items in the subscale. Since no training subjects in the Black or Hispanic groups, and very few in the control group had M-levels of six, only subjects with an M-level of 1 to 5 are included in the analysis and subscale six is not examined.

Statistical significance of the t-values is based on planned comparisons (Kirk, 1968). All comparisons are one-tail according to the expected direction. The expected direction appears in the footnote at the bottom of Table 26 for those comparisons that are significant. Eta-squared (η^2) represents the amount of explained variance and has been multiplied by 100 to convert it to a percentage; in most cases only explained variances of at least 2% are reported. Also, it is pointed out that η^2 is not independent across comparisons since the contrasts are not orthogonal. Nevertheless, it is presented because of the decrease in sample size as fewer students are included in the analysis at higher M-levels.

Results of items with an M-demand of one show a training effect for Black students on both culture-loaded and non-culture loaded items. None of the other contrasts is significant The explained variance on culture-loaded items is, however, larger than on non-culture loaded items. The interaction term which indicates a 20% greater gain for Blacks over Anglos was also not significant and accounted for only 2% of the explained variance.

Two-M items showed the same pattern of significance for culture-loaded and non-culture loaded items. For culture-loaded items neither the training effect for the Blacks nor the interaction was significant. The explained variance, however, is noticibly larger between control Blacks and Anglos on culture-loaded items. In comparison to the percent explained in the training groups and for non-culture loaded items, this suggests that group differences were greater on culture-loaded and that the differences was reduced through training.

Results of items with an M-demand of four are exactly consistent with all culture-loading expectations. That is, greater group differences occur on culture-loaded items, but training eliminated the difference. The same pattern, however, did not occur on non-culture loaded items, and Black students showed significantly greater gains due to training than Anglo students. Of note is that training had the effect of eliminating the original 18.5% variance accounted for by ethnic group differences on these items.

Results for Five-M items are also consistent with the culture-loaded expectation. While between group differences were not significant on culture-loaded items (probably because of small N's), the significant training effect for Blacks (23% variance accounted for) and significant interaction (10% explained variance) indicate that a culture-loaded interpretation is supported. In contrast, non-culture loaded items showed a significant training effect for Anglo students and also accounted for 23% explained variance.

When subscales are combined to form a total culture-loaded and non-culture loaded score, results for all subjects (M-levels 1 thru 5) are consistent with a culture-loaded hypothesis. Significant ethnic group differences occur on culture-loaded items but not on non-culture loaded items; the difference is reduced after training. Training on culture-loaded items is effective for Blacks only, and a significant gain occurs for Blacks. On non-culture loaded items, training produced significant gains for both groups and resulted in significant group differences in performance on non-culture loaded items. Nevertheless, group differences after training accounted for 5% of the variance on culture-loaded items prior to training.

In terms of the culture-loaded hypothesis results are consistent with expectations. That is, 1) significant group differences occur on culture-loaded items, 2) this difference is reduced or virtually eliminated by training, 3) training was differentially effective for Blacks on culture-loaded items, and 4) the gains for Black students on culture-loaded items was generally greater than that of Anglo students.

Hispanics: Percent passing items identified as culture-loaded and non-culture loaded for Hispanics is shown in Table 27. The percent passing each subscale is based on students with an M-level at least equal to the M-demand of the item.

Statistical analysis of the data for Hispanics is shown in Table 28. Total performance is given at the bottom of Table 28 and is computed on students with M-levels between 1 and 5 inclusive.

The pattern of performance for Hispanics is consistent with a culture-loaded hypothesis as follows: 1) significant group differences on culture loaded but not non-culture loaded items, 2) elimination of group differences on culture-loaded items due to training, 3) significant training effects for Hispanics but not for Anglos on culture-loaded items, and 4) significantly greater gains for Hispanics over Anglos on culture-loaded items.

In general, the results follow the expected pattern. Only items with an M-demand of one, however, follow the pattern completely. At least part of the hypothesis was supported on the remaining subscales. In some cases

Table 27

Percent Passing Culture-Loaded and Non-Culture Loaded Items for Hispanic and Anglo Training and Control Groups

		Culture Loaded Items				Non-Culture Loaded Items			
M-Demand		Control		Training		Control		Training	
		Hispanic	Anglo	Hispanic	Anglo	Hispanic	Anglo	Hispanic	Anglo
	$\bar{X}$	.72	.97	.92	.92	.95	.98	.98	.99
ONE	SD	.46	.17	.27	.28	.09	.06	.05	.02
(1≤M-Level≤5)	N	39	35	26	25	39	35	26	25
	$\bar{X}$	.67	.92	.78	.89	.78	.91	.75	1.00
TWO	SD	.29	.12	.26	.12	.28	.19	.31	0
(2≤M-Level≤5)	N	34	31	24	23	34	31	24	23
	$\bar{X}$	.53	.76	.69	.82	.64	.69	.73	.80
THREE	SD	.30	.22	.33	.18	.31	.29	.31	.20
(3≤M-Level≤5)	N	27	30	20	21	27	30	20	21
	$\bar{X}$	.36	.63	.64	.73	.58	.63	.70	.77
FOUR	SD	.25	.23	.27	.24	.24	.20	.26	.18
(4≤M-Level≤5)	N	21	23	14	14	21	21	14	14
	$\bar{X}$	.17	.42	.71	.67	.18	.19	.41	.46
FIVE	SD	.35	.47	.27	.43	.20	.14	.19	.23
(M-Level = 5)	N	9	12	7	9	9	12	7	9
	$\bar{X}$	.50	.73	.66	.77	.58	.64	.66	.75
TOTAL	SD	.25	.19	.26	.12	.14	.11	.15	.14
	N	39	35	26	25	39	35	26	25

Table 28

Group Comparisons on Culture Loaded and Non-Culture Loaded Raven Subscales for Hispanic and Anglo Training and Control Groups

	Culture Loaded				Non-Culture Loaded			
	$\hat{\psi}$	SE $\hat{\psi}$	t	n^2	$\hat{\psi}$	SE $\hat{\psi}$	t	n^2
M-Demand=1 (1≤M-Level≤5)								
Control: Hispanic-Anglo	-.25	.0750	-3.38*	8.5	-.02	.0152	-1.54	1.9
Training: Hispanic-Anglo	.00	.0902	0.03	-	-.02	.0183	-1.05	-
Hispanic: Training-Control	.21	.0816	2.52**	5.0	.02	.0165	1.40	-
Anglo: Training-Control	-.05	.0844	-0.61	-	.02	.0171	1.10	-
Interaction	.26	.1173	2.19***	3.6	.004	.0237	0.18	-
M-Demand=2 (2≤M-Level≤5)								
Control: Hispanic-Anglo	-.25	.0537	-4.66*	16.3	-.13	.0586	2.21*	3.9
Training: Hispanic-Anglo	-.11	.0630	-1.67	2.1	-.25	.0689	3.63*	10.4
Hispanic: Training-Control	.11	.0576	1.92	2.7	-.03	.0630	-0.545	-
Anglo: Training-Control	-.03	.0595	-0.58	-	.09	.0650	1.32	-
Interaction	.15	.0828	1.75***	2.3	-.12	.0905	-1.33	-
M-Demand=3 (3≤M-Level≤5)								
Control: Hispanic-Anglo	-.22	.0700	-3.21*	9.5	-.06	.0750	-0.75	-
Training: Hispanic-Anglo	-.12	.0825	-1.46	1.9	-.07	.0884	-0.79	-
Hispanic: Training-Control	.16	.0779	2.06**	3.8	.09	.0835	1.14	-
Anglo: Training-Control	.06	.0751	0.82	-	.11	.0805	1.33	-
Interaction	.10	.1082	0.91	-	-.01	.1160	-0.12	-
M-Demand=4 (4≤M-Level≤5)								
Control: Hispanic-Anglo	-.27	.0746	-3.67*	14.6	-.05	.0670	-0.70	-
Training: Hispanic-Anglo	-.08	.0934	-0.89	-	-.07	.0840	-0.85	-
Hispanic: Training-Control	.29	.0852	3.35**	12.2	.11	.0766	1.48	2.9
Anglo: Training-Control	.10	.0837	1.44	2.3	.14	.0753	1.83	4.5
Interaction	.19	.1195	1.59	2.8	-.02	.1074	-0.23	-
M-Demand=5 (M-Level=5)								
Control: Hispanic-Anglo	-.25	.1775	-1.41	4.6	-.01	.0829	-0.08	-
Training: Hispanic-Anglo	.05	.2028	0.24	-	-.05	.0947	-0.50	-
Hispanic: Training-Control	.55	.2028	2.70**	16.9	.23	.0947	2.43**	11.9
Anglo: Training-Control	.25	.1775	1.41	4.6	.27	.0829	3.27	21.4
Interaction	.30	.2695	1.10	2.8	-.04	.1259	-0.32	-
All Items (1≤M-Level≤5)								
Control: Hispanic-Anglo	-.23	.6503	-4.58*	13.7	-.06	.0291	-2.23*	3.3
Training: Hispanic-Anglo	-.11	.0605	-1.86	2.3	-.09	.0351	-2.61*	4.6
Hispanic: Training-Control	.16	.0547	2.99**	5.8	.08	.0317	2.50**	4.2
Anglo: Training-Control	.05	.0565	0.81	-	.11	.0328	3.23**	7.0
Interaction	.12	.0786	1.50	1.5	-.03	.0456	-0.59	-

* Significant at $p < .025$, $t < -1.96$

** Significant at $p < .025$, $t > 1.96$

*** Significant at $p < .05$, $t > 1.645$

non-significance was probably related to small sample size, however, the explained variance followed the expected pattern. For example, on items with an M-demand of five, significance occurred in only one instance (training effect for Hispanics). However, the 4.6% variance explained by ethnic group difference for control group students and the lack of variance explained by ethnic group difference with training group students follows the expected outcome.

Culture-loaded items with an M-demand of two produced ethnic group differences in the control groups as did non-culture loaded items. The amount of explained variance, however, was 16% on culture-loaded items in comparison to 4% on non-culture loaded items. Additionally, training reduced the ethnic group difference on culture-loaded items but not on non-culture loaded items. The effect of training was 14% higher on culture-loaded items than on non-culture loaded items, but was not significant. Finally, and perhaps more important, is that the gains for Hispanics on culture-loaded items was statistically greater than the gains for Anglos. Even though the observed treatment effect was not significant, it is clear that the overall effect on culture-loaded items was consistent with expectation. The interaction contrast for non-culture loaded items was not significant.

On items with an M-demand of three, significant group differences occurred on culture-loaded items but not on non-culture loaded items. The difference did not hold up after training, and training was significant for Hispanic but not for Anglo students. Only the non-significant interaction failed to support the culture-loaded hypothesis for this subscale.

Four-M items followed the same pattern as items with an M-demand of three. Again, only the interaction was not as expected; all other contrasts were consistent with expectation.

Five-M items showed significant Training effect for Hispanics on both types of items. On culture-loaded items the variance explained was 17% in comparison to 12% for non-culture loaded items.

Finally, the results for all students across all items were consistent with expectation except for the interaction. Overall, greater group difference occurred on culture-loaded items and explained 14% of the variance in contrast with 3% explained on non-culture loaded items. Training had the effect of removing ethnic group differences on culture-loaded items but not on non-culture loaded items.

Summary of Results on Culture-Loading

Identification of items as culture-loaded was based on outcomes or results considered to be consistent with the culture-loaded hypothesis. These are: 1) significant ethnic group differences on culture-loaded items; 2) elimination or reduction of ethnic group differences on culture-loaded items, after training; 3) significant training effects for minority students but not Anglo students on culture-loaded items; and, 4) greater training effects on culture-loaded items for minorities than for Anglos.

In general these outcomes were supported on all the Raven subscales, and when subscales were combined to form culture-loaded and non-culture loaded totals. The outcome that occurred with the least frequency was a significant interaction effect, i.e., greater gains between minority training and control students than for Anglo training and control students. In total, the results were consistent with an expectation derived from a culture-loaded explanation.

The results are also of significance because of the fact that the outcomes were predicted for certain items and because more than one outcome (i.e., pattern) was correctly predicted. In cases where all expectations did not occur there was generally just one outcome that was not as predicted. Admittedly, the procedure for selecting the items will tend to produce greater group differences on the so-called culture-loaded items, since, after all, they were the items which showed a 25% or more difference at a given M-level in the first place. On the other hand, not all such items necessarily showed the greatest group differences, nor did items not identified by this criteria not show group differences. The more significant fact is that the pattern in the five comparisons was generally supported as predicted, especially greater gains (i.e., effects for particular groups but not others) suggests the procedure is sound. It is not group differences alone, but the consistent overall pattern that is significant.

The results indicate that in most instances items were correctly identified in which training was likely to result in improved performance for minority students. In some cases Anglo performance was affected positively too. Overall, the effect was to bring the performance of minority and majority groups closer together.

Finally, several comments are in order regarding the results presented. First, many statistical tests were performed and one would expect a certain percent to be significant by chance factors alone. It is pointed out, however, that the alpha level was controlled for each family of tests. This allows for some control of the Type I error rate. Perhaps more important is the pattern of significance. For example, had all statistical tests been performed at .05, then roughly 5% would be significant simply by chance. The important thing to observe is that significance occurred on precisely those outcomes predicted to be significant, and in the predicted direction. Too, outcomes predicted to not be significant generally were not. Thus, significance did not occur on a random basis. It is the consideration of both the control of the alpha level and the pattern of significant outcomes that support the results. At the same time, the question is not really one of statistical significance in the first place. In the final analysis, it is the overall consistency in which the outcomes hold together.

Second, the reader is reminded of at least two sources of error that mitigate against the consistency of the observed outcomes. One is that it relied on the identification of an item's M-demand by external sources. To the extent that there was error in this process, it surely effected the outcomes. Another is the identification of subjects' M-level. The FIT itself could have introduced a source of error. The finding of a certain percentage of minority and Anglo students with M-levels less than or equal to zero, and the realization that training was effective for these students, suggest that some measure of error occurred in the respect.

As a final note, it is argued that children did not acquire the particular skills necessary to take the Raven test, nor were they taught. It is the authors' position, though not tested, that children have the required underlying test taking skills and that we simply provided them with the chance to learn how to apply them in this particular situation. And, while teachers noted changes in childrens' approach to many classroom tasks, we have no evidence as to the effects of the training outside of the Raven test.

In this study, only eight hours of training was provided. During the training, we did not show the children how to solve a particular task. We simply pointed out the errors they were making, and encouraged them to develop alternative strategies to the extent that they became aware that the one they were applying was either not sufficient or wrong. Often children assumed their answer was correct and did not attempt to check their answers or look for alternative solutions. Had the children not had the skills in the first place, it is unlikely, that the training could have provided them in such a short period of time. It is also unlikely that training could produce marked jumps in developmental abilities.

Finally, a few words are necessary concerning Hispanic students in the study. Many of the students were in bilingual education programs, spoke limited or no English and thus had to receive training in Spanish. However, this proved to be difficult because of the need for "on the spot translations" as well as the use of a particular vocabulary. The training required words not commonly used by children or the trainers since many of them had not been "educated" in a Spanish school system. While every attempt was made to focus on communication, there was undoubtedly error or ambiguity introduced due to language. An attempt was originally made to gather language proficiency data in English and Spanish, but was discontinued because of cost and time constraints. Many of the children, however, spoke only Spanish and several more were classified as limited English speakers. The exact effect of this source of uncontrolled error is not known. It may have contributed to performance differences between Hispanics, Blacks, and Anglos, and it undoubtedly contributed to poorer success in the effect of training.

CONCLUSION

Overall, it appears that the procedure, while admittedly not clear cut, did produce results which suggest that there is a form of cultural test bias in that culture-loaded items could be identified which produced lower scores for particular groups of children and which were differentially affected by training as predicted. The fact that specific items could be identified in advance, and that training was differentially effective sheds new light on test performance in general. The results should provide a better understanding of, and insights into what tests are measuring irrespective of ethnic background.

Important too, is the fact that a particular cognitive developmental theory was shown to be useful and related to psychometric test performance. It is significant that evidence was provided for a specific source of differences in test performance. That is, that test performance error, whether one calls it lack of validity or bias, is in part due simply to test taking skills. In all, the training was really an exercise in providing test taking skills to a particular group of children who, prior to training, did not know how to apply them to the particular paper and pencil tests used.

One cannot discuss the question of culture-loading and test bias without the issue of group differences arising. The results in this study indicate that while group differences on the Raven appear to be related to group differences in developmental level, this is not entirely the case. That is, even though minority students tend to score lower on developmental measures and to show an overall proportion of students obtaining a lower

M-level, two results suggest that it is test taking skills which affect paper-pencil tests in general. For example, when performance of subjects within M-levels is examined, ethnic group differences in performance are observed in the control but not the training groups. Training group students showed virtually the same performance with M-levels, with minority students often scoring higher. In addition, there were not significant ethnic group differences between training group subjects in spite of the fact that more minority subjects had lower M-levels. In short it is unlikely that training of such short duration could have an effect on cognitive developmental level, or have the differential effect in favor of minority students unless there was a learned skill involved. It is also more likely that the M-level performance of minority students was also depressed. The higher performance of minority students with M-levels in the training group is consistent with this interpretation. In all, it is interpreted that test taking skills are a major source of variation, that these skills are learned, and that they can be overcome through exposure to the specific requirements of the test.

These results suggest that the question of test bias is not necessarily in the test itself. It is in the overall testing procedure, most important of which is the assumption that all children approach and solve a task according to the way in which the test publisher wishes. The indiscriminant use of tests without awareness or consideration of this factor will result in errors in validity as well as bias. The conscientious test publisher, as well as test user should be well aware of these problems. This study provides evidence that at least this one source of error has not been considered, that is not commonly taken into account by test developers or users, and that no amount of statistical or psychometric validation is likely to account for this error. The use of cognitive or some other theory of mental functioning is sorely needed in development and validation of tests. The researchers hope that this study will provide incentive as well as direction for future fair uses of tests.

References

Bachelder, B. L., & Denny, M. R. A theory of intelligence: I. Span and the complexity of stimulus control. Intelligence, 1977a, 1, 127-150.

Bachelder, B. L., & Denny, M. R. A theory of intelligence: II. The role of span in a variety of intelligence tasks. Intelligence, 1977b, 1, 237-256.

Bereiter, C. & Scardamalia, M. Pascual-Leone's M-construct as a link between cognitive-developmental and psychometric concepts of intelligence. Intelligence, 1979, 3, 41-63.

Buss, Allan R. The relationship between the psychological environment and the development of individual differences in abilities. Intelligence, 1977, 2, 192-207.

Case, R. Validation of a neo-Piagetian mental capacity construct. Journal of Experimental Child Psychology, 1972, 14, 287-302

Case, R. Mental stategies, mental capacity, and instruction: a neo-Piagetian investigation. Journal of Experimental Child Psychology, 1974, 18, 382-397.

Case, R. Social class differences in intelectual development: A neo-Piagetian investigation. Canadian Journal of Behavioral Science, 1975, 7, 244-261.

Case, R., & Globerson, T. Field independence and central computing space. Child Development, 1974, 45, 772-778.

Case, R., & Pascual-Leone, J. Failure of conservation training of disadvantaged Black teenagers: A neo-Piagetian interpretation. Perceptual Skills and Motor Development, 1975, 40, 545-546.

Castañeda, A. Cultural democracy and the educational needs of Mexican-American children. In P. L. Jones (Ed.), Mainstreaming and the minority child. Mineapolis, Mn.: The Coucil for Exceptional Children Publication, 1976.

Cole, M. An ethnographic psychology of cognition. In R. Brislin, S. Bochner, & W. Lonner (Eds.), Cross-Cultural Perspectives on Learning. New York: Halstead Press, 1975.

Cole, M., & Bruner, J. S. Cultural differences and inferences about psychological processing. American Psychologist, 1971, 26.

Cole, M., & Gay, J. Culture and memory. American Anthropologists, 1972, 74, 1066-1084.

Cole, M., Gay, J., Glich, J. A., & Sharp, D. W. The Cultural Context of Learning and Thinking. New York: Basic Books, 1971.

Das, J. P. Structure of cognitive abilities: Evidence for simultaneous and successive processing. Journal of Educational Psychology, 1973a, 65, 103-108.

Das, J. P. Cultural deprivation and cognitive competence. In N. R. Ellis (Ed.), International Review of Research in Mental Retardation, Vol. 6, New York: Academic Press, 1973b.

Das, J. P., Kirby, J., & Jarman, R. F. Simultaneous and successive syntheses: An alternative model for cognitive abilities. Psychological Bulletin, 1975, 82, 87-103.

De Avila, E. A. *Cartoon Conservation Scales (CCS), Level I and II*. Corte Madera, Ca.: Linguametrics Group, 1977.

De Avila, E. A., & Duncan, S. E. *The construct validity of several measures of cognitive style for language minority children*. Austin, Tx.: Southwest Educational Development Laboratory, 1979.

De Avila, E. A., & Havassy, B. E. The testing of minority children: A neo-Piagetian alternative. *Today's Education, Jounal of the National Education Association*, November/December 1974, 72-75.

De Avila, E. A., Havassy, B. E., & Pascual-Leone, J. Mexican-American school children: A neo-Piagetian analysis. Washington, D.C.: Georgetown University Press, 1976.

DeVries, R. Relationships among Piagetian, IQ and achievement assessments. Chicago, Il.: ERIC document reproduction service, ED-088-593, March 1973.

Duncan, S. E., & De Avila, E. A. *Relative linguistic proficiency and field dependence/independence: Some findings on the linguistic heterogeneity and cognitive style of bilingual children*. Paper presented at the 13th Annual Convention of TESOL, Colloquium on ESL in Bilingual Education: Implementation and Other Practical Considerations, Boston, Mass., February 27-28, 1979.

Dunn, O. J. Multiple comparisons among means. *Journal of the American Statistical Association*. 1961, 56, 52-64.

Feurenstein, R. *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore: University Park Press, 1979.

Hathaway, W. E., Jr. The degree and nature of the relations between traditional psychometric and Piagetian developmental measures of mental development. Paper presented at the annual meeting of the American Educational Research Association, March 1973.

Hunt, E., Quote the Raven? Nevermore! In L. Gredd (Ed.) *Knowledge Cognition*. Erlbaum Associates, 1974.

Jensen, A. R. Level I and Level II abilities in three ethnic groups. *American Educational Research Journal*, 1973, *4*, 263-276.

Jensen, A. R. Interaction of Level I and Level II abilities with race and socioeconomic status. *Journal of Educational Psychology*, 1974a, *66*, 99-111.

Jensen, A. R. How biased are culture-loaded tests? *Genetic Psychology Monographs*, 1974b, *90*, 185-244.

Jensen, A. R. The nature of intelligence and its relation to learning. Paper presented at the Theodore A. Fink Memorial Lecture, University of Melbourne, in September, 1977.

Jensen, A. R. *g*: Outmoded theory or unconquered frontier. *Creative Science & Technology*, 1979, Vol, II, No. 3, 16-29.

Jensen, A. R. Definitions and Criteria of Test Bias, chapter from *Test Bias* (in press), 1980.

Kerlinger, F. N., & Pedhauzer, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart & Wiston, Inc, 1973.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Brooks/Cole, Belmont, CA., 1968.

Kohlberg, L., & DeVries, R. Relations between Piaget and psychometric assessments of intelligence. Paper presented at the Conference on the Natural Curriculum of the Child, Urbana, Illinois, 1969.

Laosa, L. M. Maternal teaching strategies and field dependence independence cognitive styles in Chicano families. Education Testing Service Research Bulleting RB-78-12, September 1978.

Laosa, L. M., & De Avila, E. A. Development of cognitive styles among Chicanos in traditional and dualistic communities. International Journal of Psychology, 1979, 14, 91-98.

Luria, A. R. Human Brain and Psychological Process. New York: Harper & Row, 1966.

Marascuilo, L.M. Statistical methods for behavioral science research. New York: McGraw-Hill, 1971.

Marascuilo, L. M., & McSweeney, M. Nonparametric and distribution free methods for the social sciences. Monterey, Ca.: Brook/Cole Publishing Co., 1977.

Marascuilo, L. M. & Levin, J. R. The Simultaneous Investigation of Two-Factor Analysis of Variance Designs. American Educational Research Journal, Winter 1976, Vol. 13, 1, 61-65.

Pascual-Leone, J. Cognitive development and cognitive style: A general psychological integration. Unpublished doctoral dissertation, University of Geneva, 1969.

Pascual-Leone, J. A mathematical model for the transition rule in Piaget's developmental stages. Acta Psychologica, 1970, 32, 301-345.

Pascual-Leone, J. A theory of cosntructive operators, a neo-Piagetian model of conservation and the problem of horizontal decaloges. Extended version of a paper presented at the Canadian Psychological Association meeting, 1972.

Piaget, J., & Inhelder, B. La representation de l'espace chez le'enfant. Paris: Presses Universitaires de France, 1948.

Ramírez, M., & Castañeda, A. Cultural Democracy, Bicognitive Development, and Education. New York: Academic Press, 1974.

Ramírez, M., Castañeda, A., & Herold, P. L. The relationship of acculturation to cognitive style among Mexican-Americans. Journal of Cross-cultural Psychology, 1974, 5, 425-433.

Ramírez M., & Price-Williams, D. Cognitive styles in children: Two Mexican communities. Interamerican Journal of Psychology, 1974, 8, 93-100.

Rohwer, W. D., Jr. Learning, race, and school success. Review of Educational Research, 1971, 41, 191-210.

Sternberg, R. J. Isolating the components of intelligence. Intelligence, 1978, 2, 117-128.

Tuddenham, R. D. The nature of intelligence. Lead off lecture presented in a lecture series entitled "Measuring the Measures of of Intelligence", University of California, San Diego, September 1972.

Ulibarri, D. M. Technical manual: Multilingual assessment scale. Multilingual Assessment Program, Stockton Unified School District, Stockton, Ca., 1974 (USOE, Title VII).

Ulibarri, D. M. Cognitive processing theory and culture-loading: A neo-Piagetian approach to test bias. Unpublished doctoral dissertation, University of California, Berkeley, 1982.

Ulibarri, D. M., & Flemming, J. S. Field dependence/independence of traditional and dualistic Chicano communities and Anglo communities. Berkeley, Ca., 1980.

Witkin, H. A. Individual differences in ease of perception of embedded figures. Journal of Personality, 1950, 19, 1-15.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A manual for the embedded figures test. Palo Alto, Ca.: Consulting Psychologists Press, Inc., 1971.

APPENDIX A

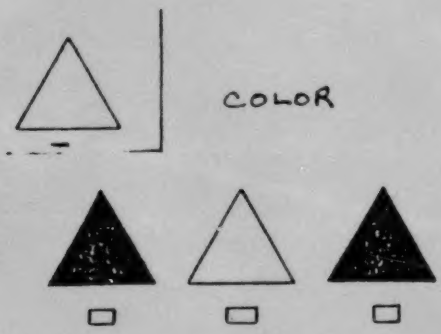
TRAINING EXERCISES

EXERCISE I

SCHOOL ____________________

NAME ____________________

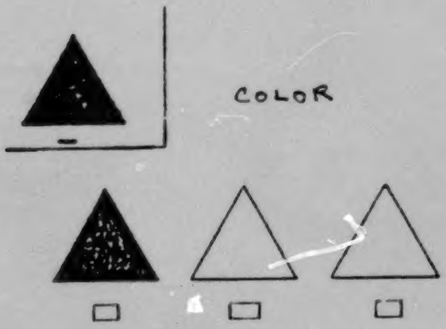
1



COLOR

☐ ☐ ☐

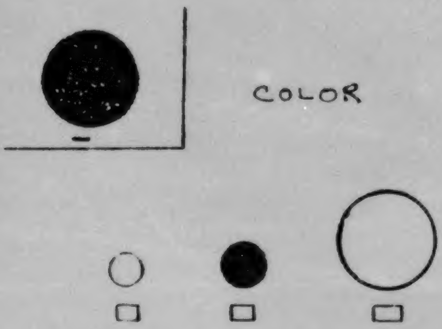
2



COLOR

☐ ☐ ☐

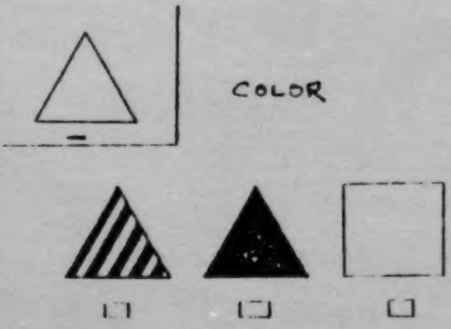
3



COLOR

☐ ☐ ☐

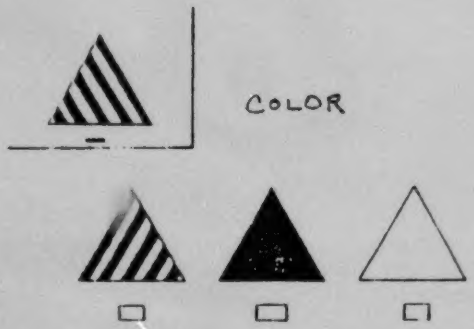
4



COLOR

☐ ☐ ☐

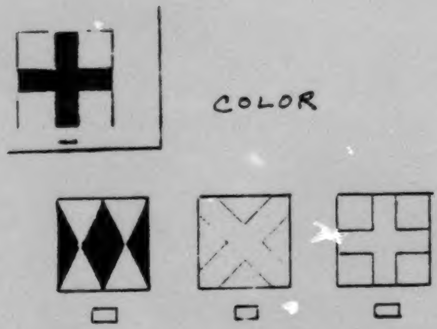
5



COLOR

☐ ☐ ☐

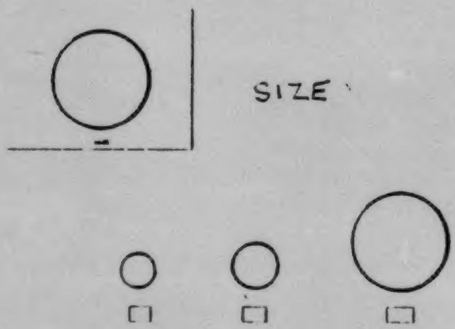
6



COLOR

☐ ☐ ☐

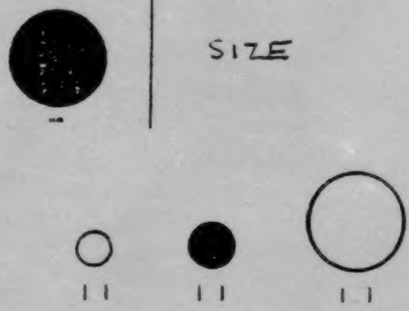
7



SIZE

☐ ☐ ☐

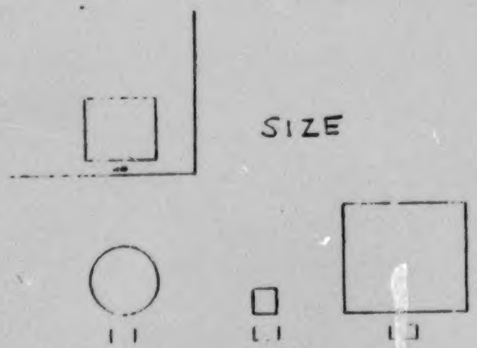
8



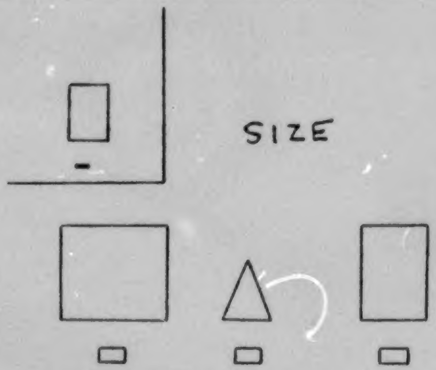
SIZE

☐ ☐ ☐

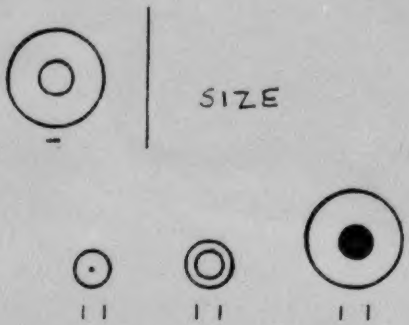
9



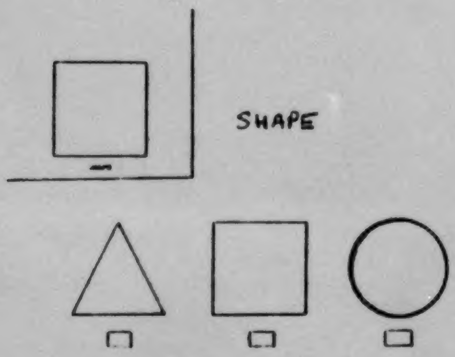
10



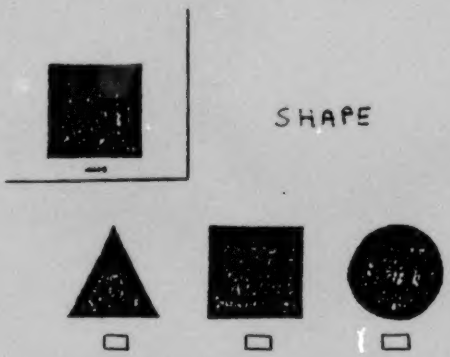
11



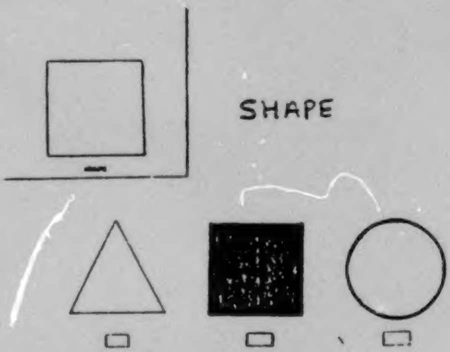
12



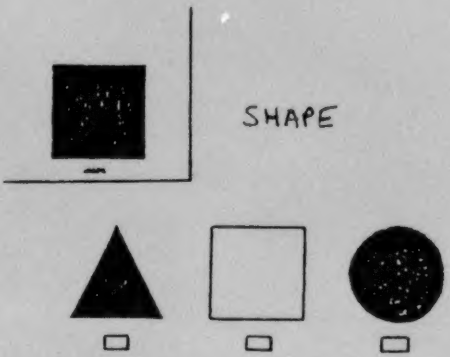
13



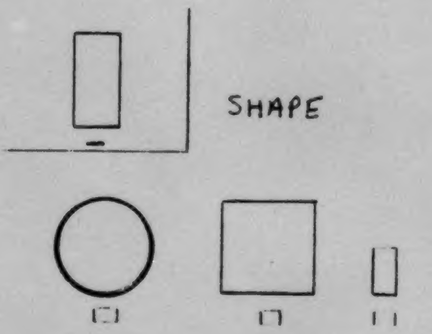
14



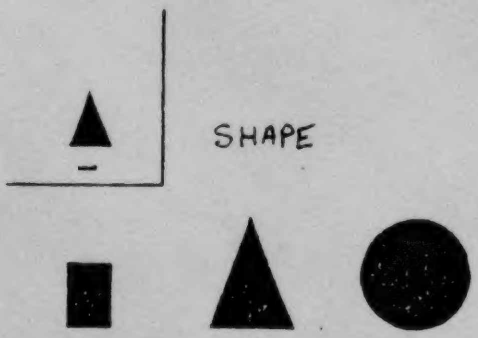
15



16



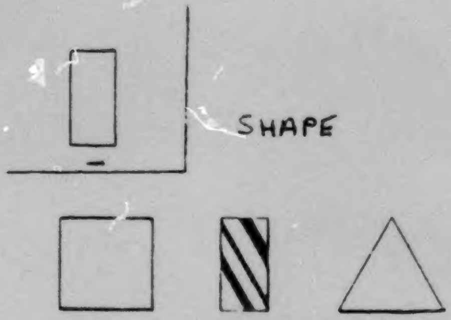
17



SHAPE

☐ ☐ ☐

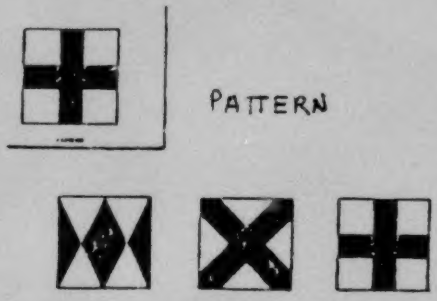
18



SHAPE

☐ ☐ ☐

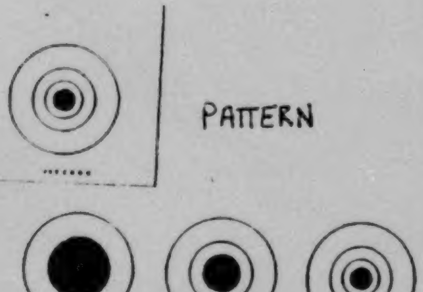
19



PATTERN

☐ ☐ ☐

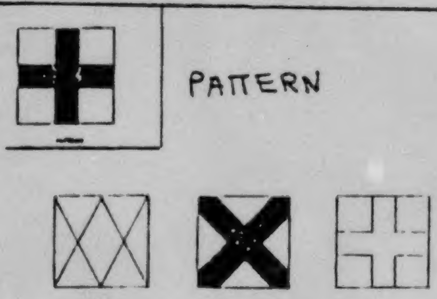
20



PATTERN

☐ ☐ ☐

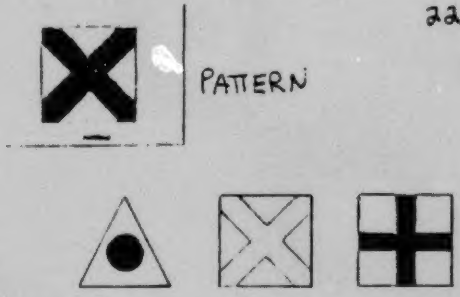
21



PATTERN

☐ ☐ ☐

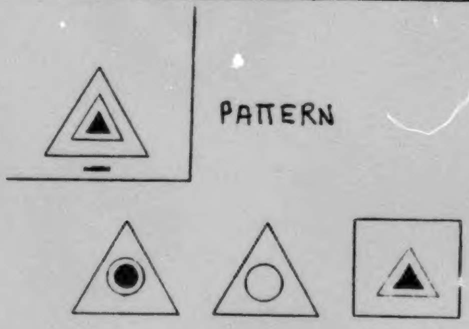
22



PATTERN

☐ ☐ ☐

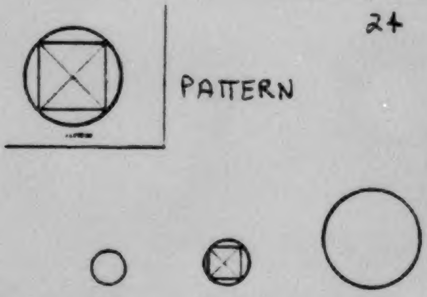
23



PATTERN

☐ ☐ ☐

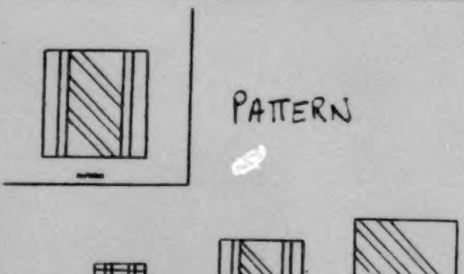
24



PATTERN

☐ ☐ ☐

25

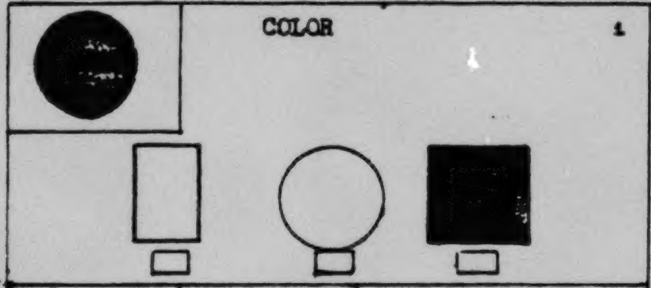


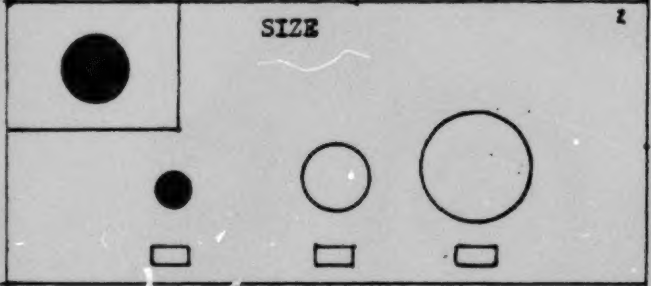
PATTERN

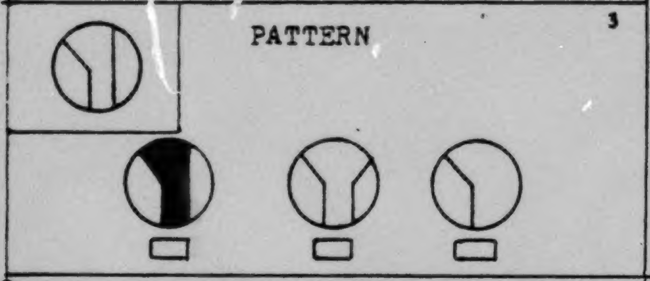
☐ ☐ ☐

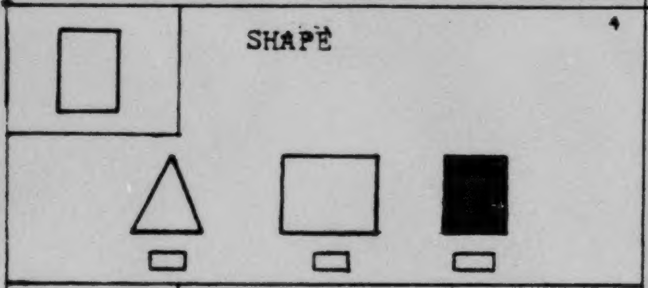
EXERCISE I: CRITERION MASTERY TEST

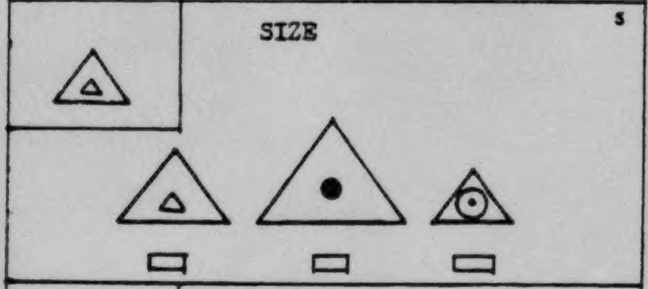
____
NAME (ID)

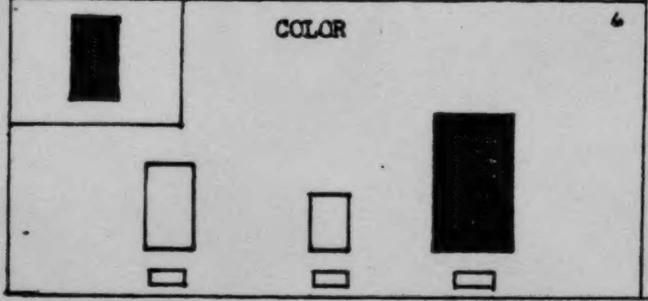

COLOR 1

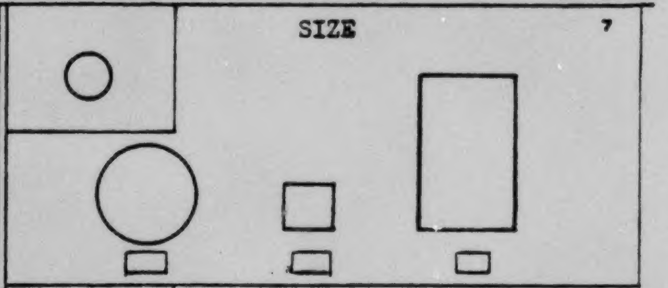

SIZE 2

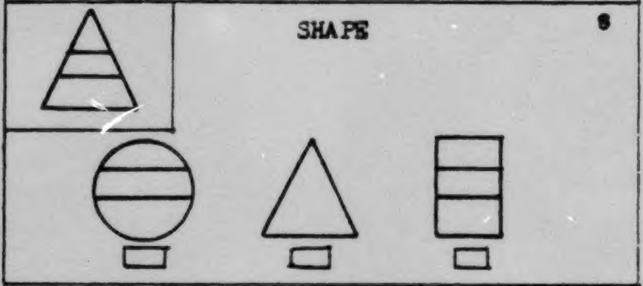

PATTERN 3

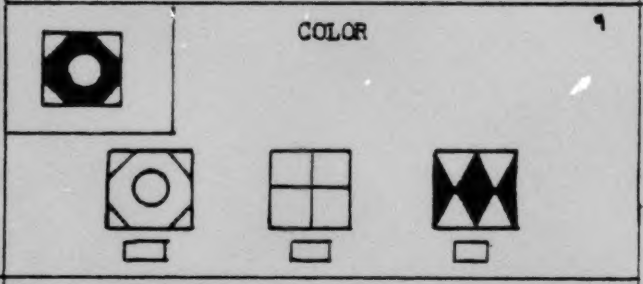

SHAPE 4

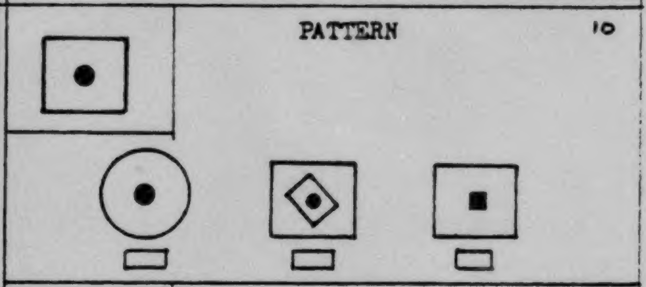

SIZE 5

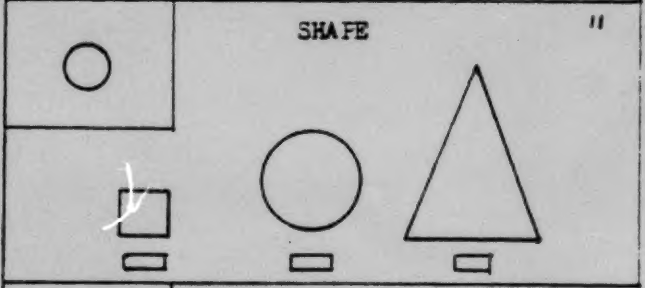

COLOR 6

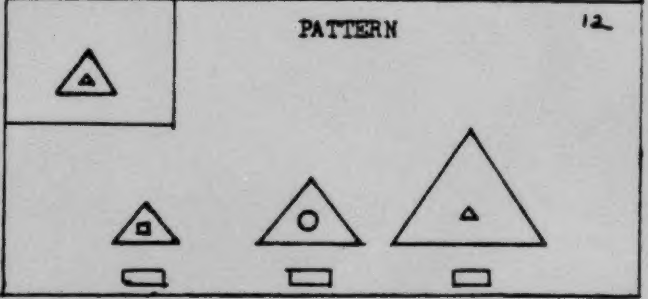

SIZE 7


SHAPE 8


COLOR 9


PATTERN 10


SHAPE 11

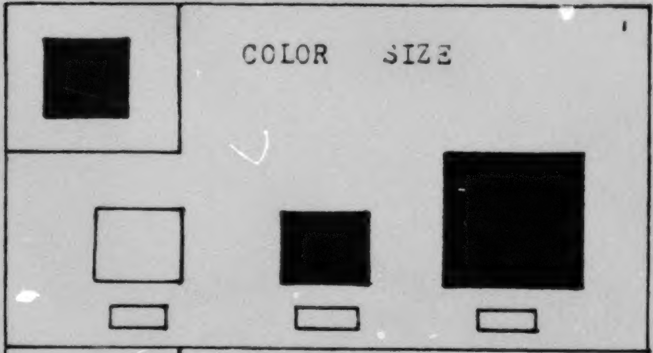

PATTERN 12

____
School

EXERCISE 2: TWO DIMENSIONS

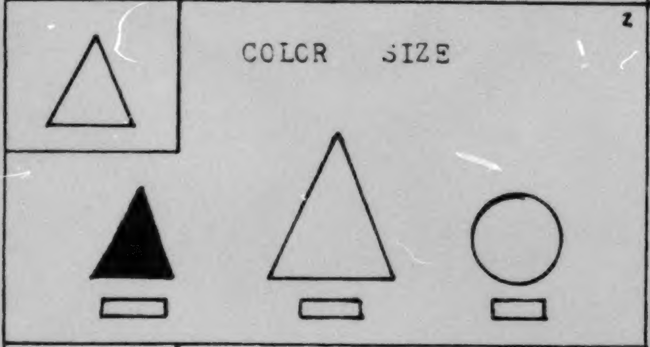
NAME (I.D.)

1



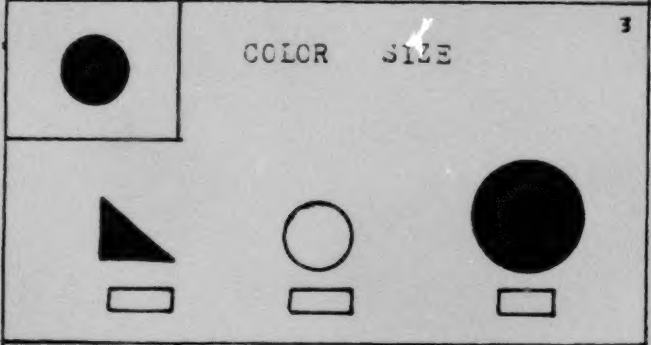
COLOR SIZE

2



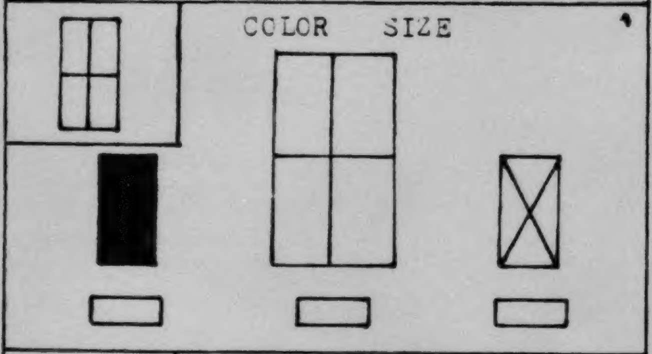
COLOR SIZE

3



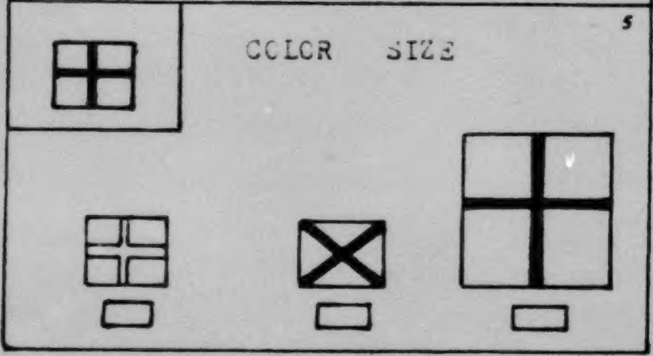
COLOR SIZE

4



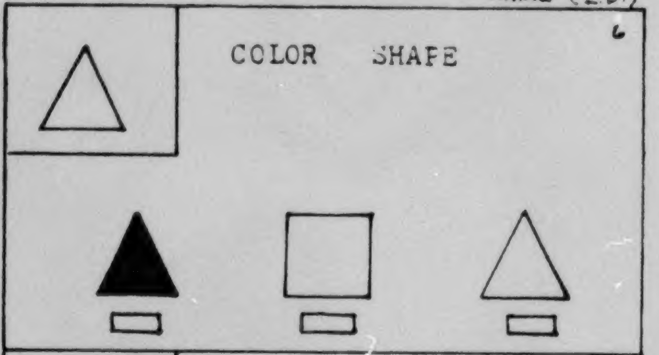
COLOR SIZE

5



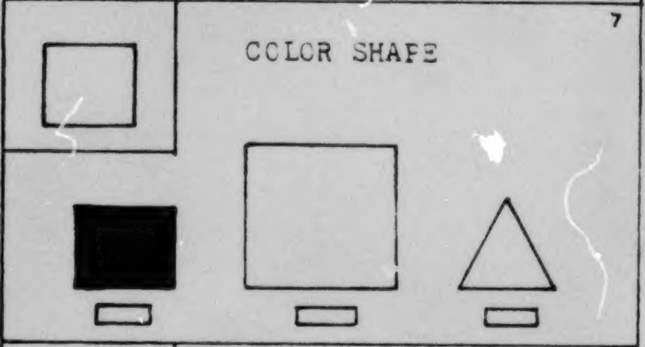
COLOR SIZE

6



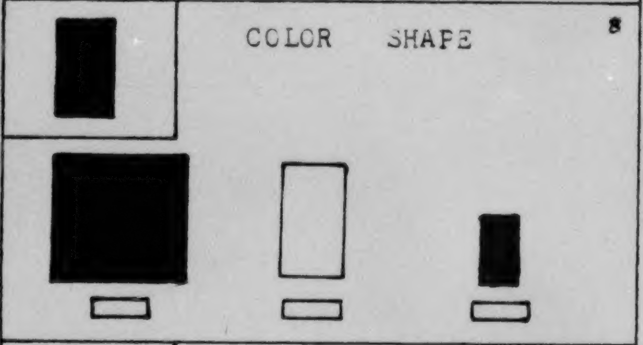
COLOR SHAPE

7



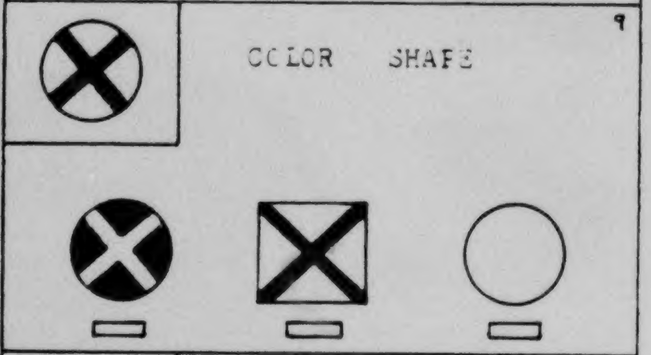
COLOR SHAPE

8



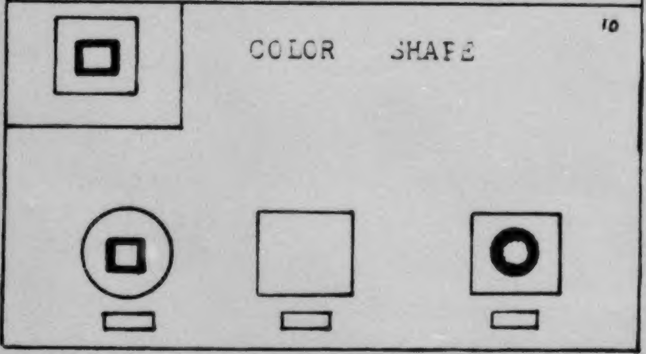
COLOR SHAPE

9



COLOR SHAPE

10

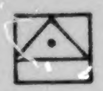


COLOR SHAPE

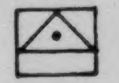
School

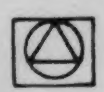
EXERCISE 2

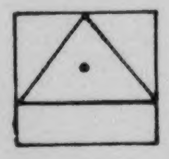
21



SIZE PATTERN

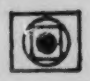




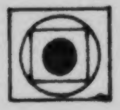


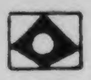
☐ ☐ ☐

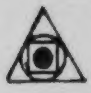
26



SHAPE PATTERN

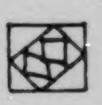




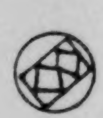


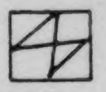
☐ ☐ ☐

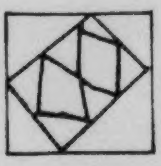
22



SIZE PATTERN



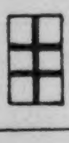




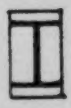
☐ ☐ ☐

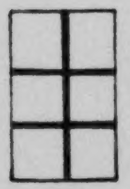
☐ ☐ ☐

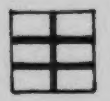
23



SIZE PATTERN



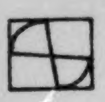




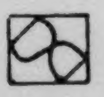
☐ ☐ ☐

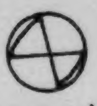
☐ ☐ ☐

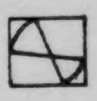
24



SHAPE PATTERN



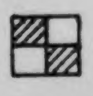




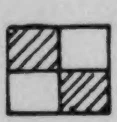
☐ ☐ ☐

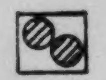
☐ ☐ ☐

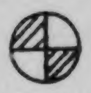
25



SHAPE PATTERN





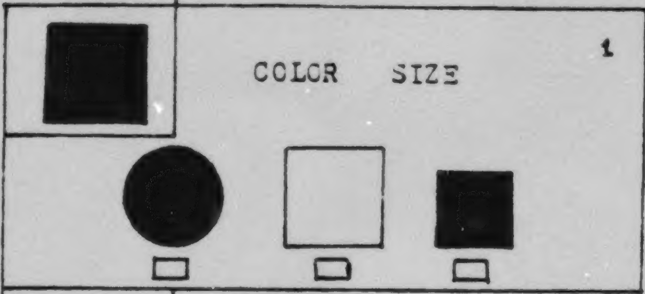


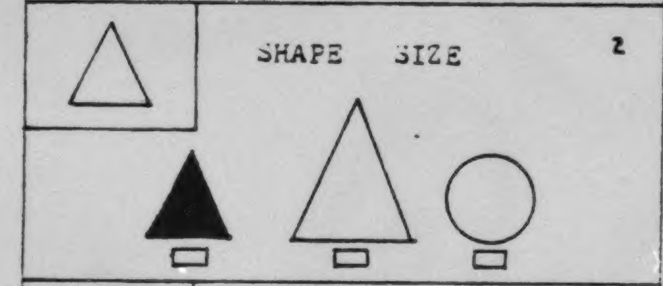
☐ ☐ ☐

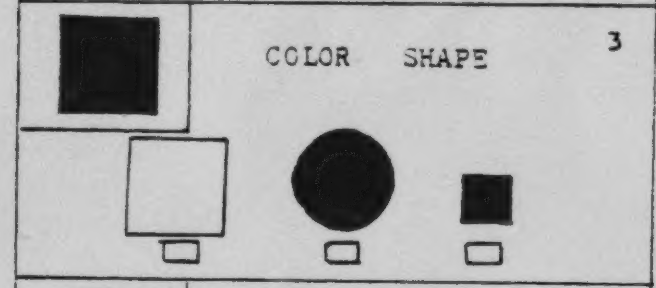
☐ ☐ ☐

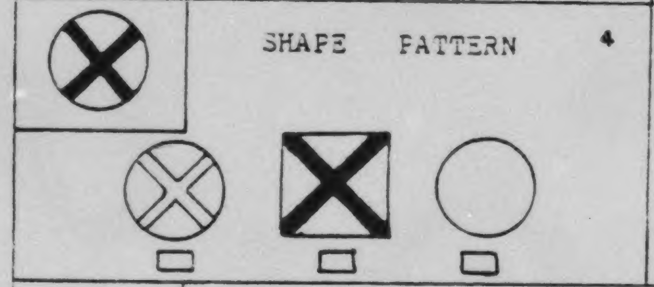
EXERCISE 2: CRITERION MASTERY TEST

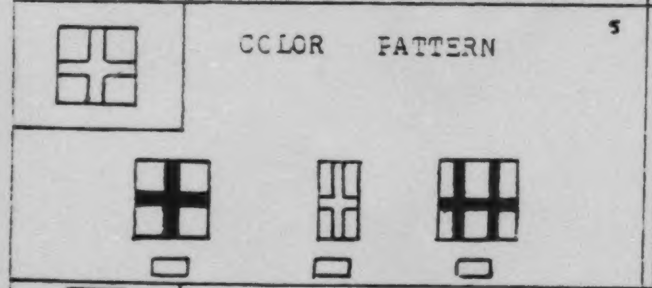
Name (ED)

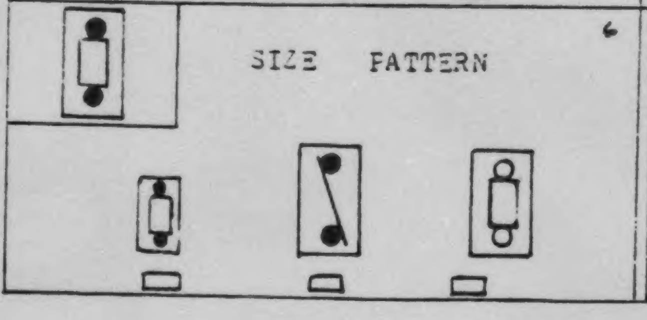

1
COLOR SIZE

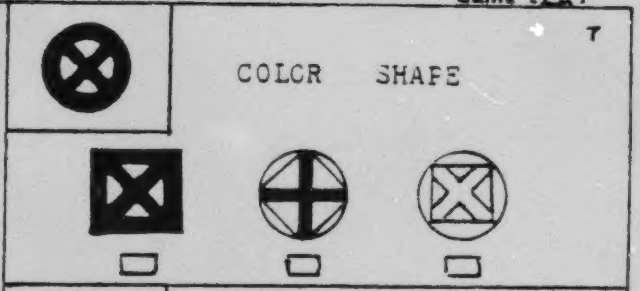

2
SHAPE SIZE

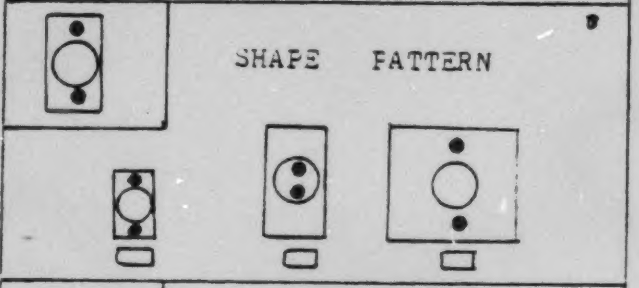

3
COLOR SHAPE

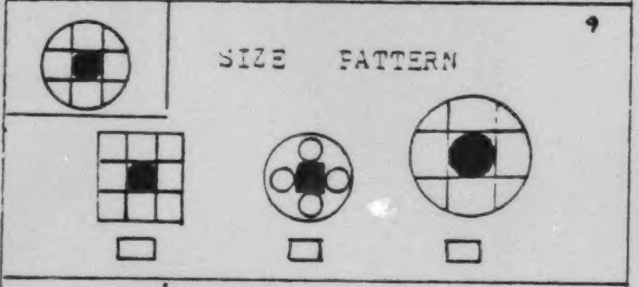

4
SHAPE PATTERN

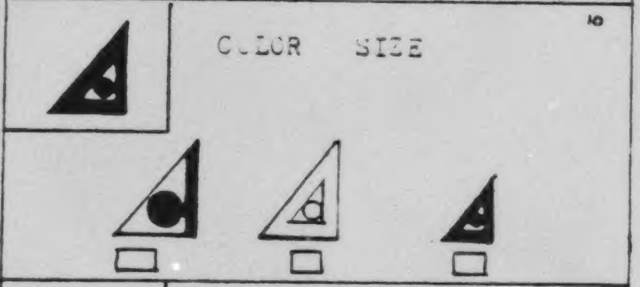

5
COLOR PATTERN

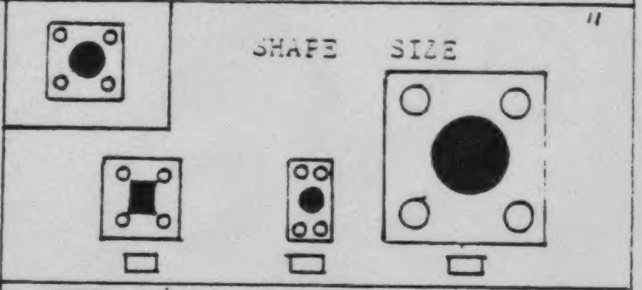

6
SIZE PATTERN

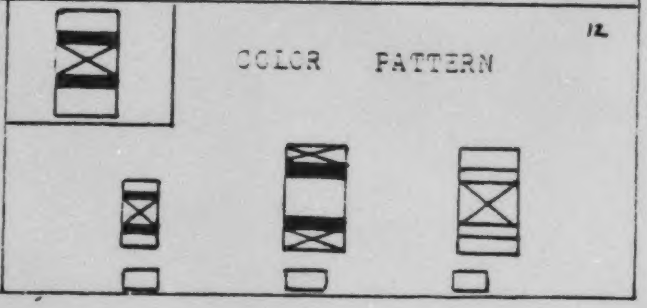

7
COLOR SHAPE


8
SHAPE PATTERN


9
SIZE PATTERN

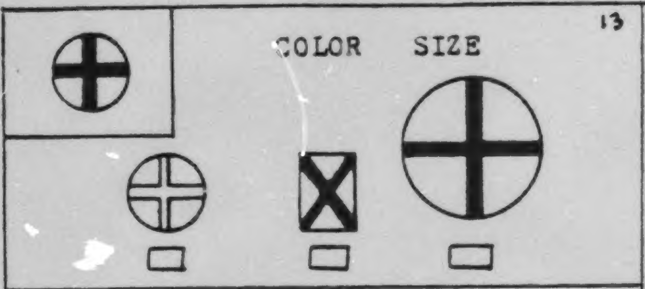

10
COLOR SIZE


11
SHAPE SIZE

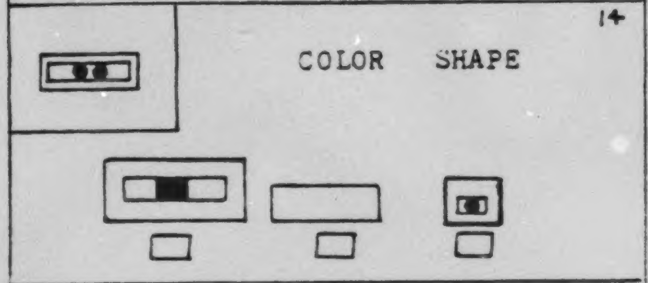

12
COLOR PATTERN

School

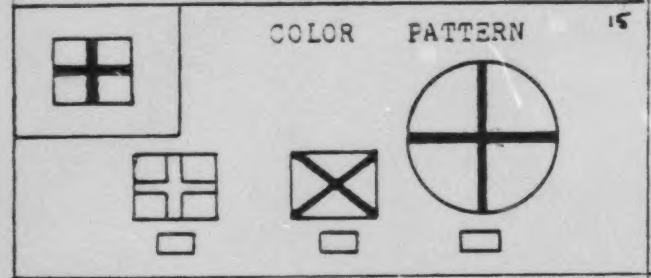
CRITERION MASTERY TEST _ PAGE 2


13
COLOR
SIZE

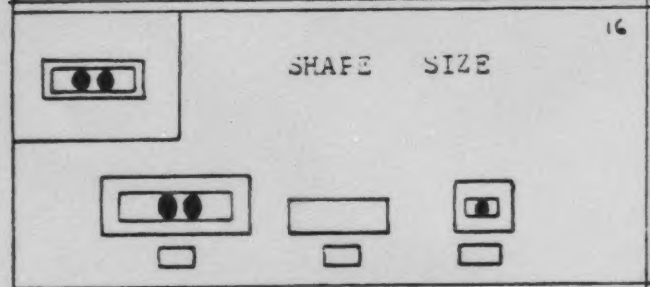
☐ ☐ ☐


14
COLOR
SHAPE

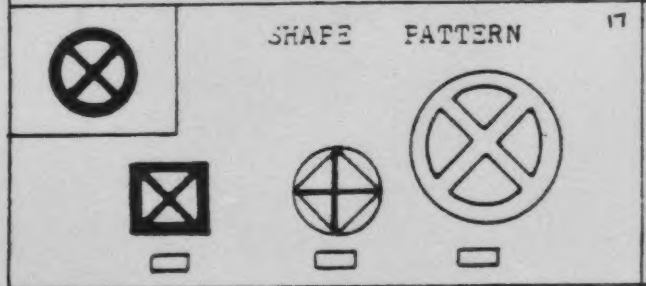
☐ ☐ ☐


15
COLOR
PATTERN

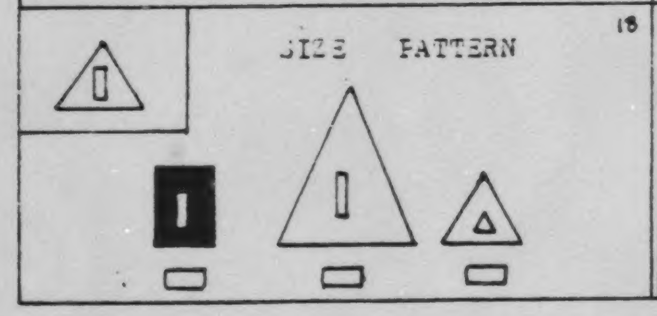
☐ ☐ ☐


16
SHAPE
SIZE

☐ ☐ ☐


17
SHAPE
PATTERN

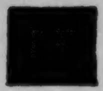
☐ ☐ ☐


18
SIZE
PATTERN

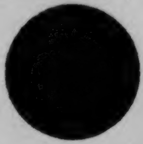
☐ ☐ ☐

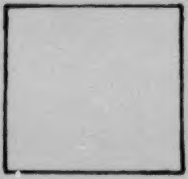
Exercise 3: Three Dimensional Training

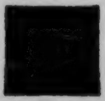
NAME

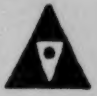


Color Size Shape

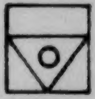

☐

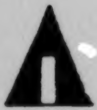

☐

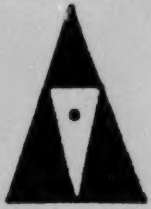

☐

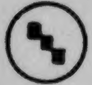


Color Size Shape

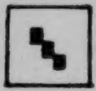

☐

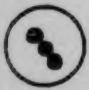

☐

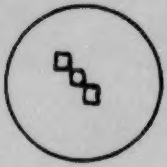

☐

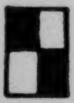


Color Size Shape

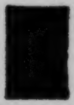

☐

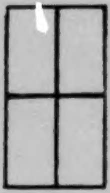

☐

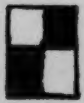

☐

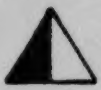


Color Size Shape

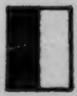

☐

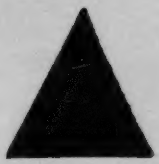

☐

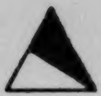

☐

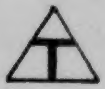


Color Size Shape

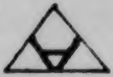

☐

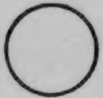

☐

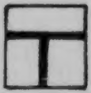

☐

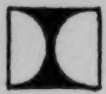


Color Size Pattern

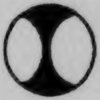

☐

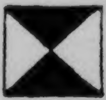

☐

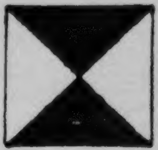

☐

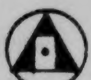


Color Size Pattern

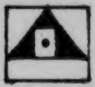

☐

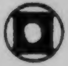

☐

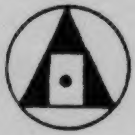

☐

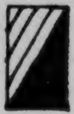


Color Size Pattern

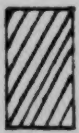

☐

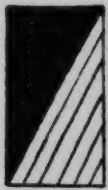

☐

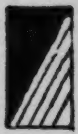

☐

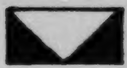


Color Size Pattern

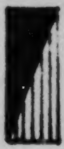

☐

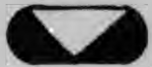

☐

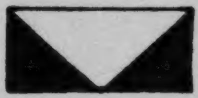

☐



Color Size Pattern


☐

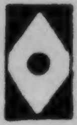

☐


☐

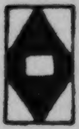
SCHOOL

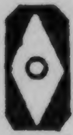
Criterion Master Test: Exercise 3

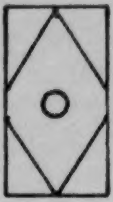
NAME

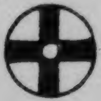


Color Size Shape

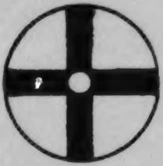


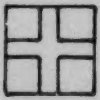


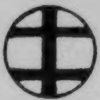


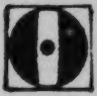


Color Size Shape

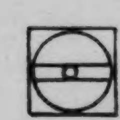


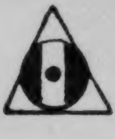


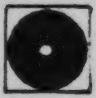


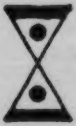


Color Size Pattern

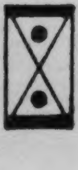


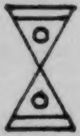


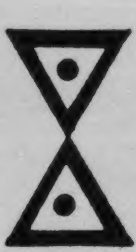




Color Size Pattern



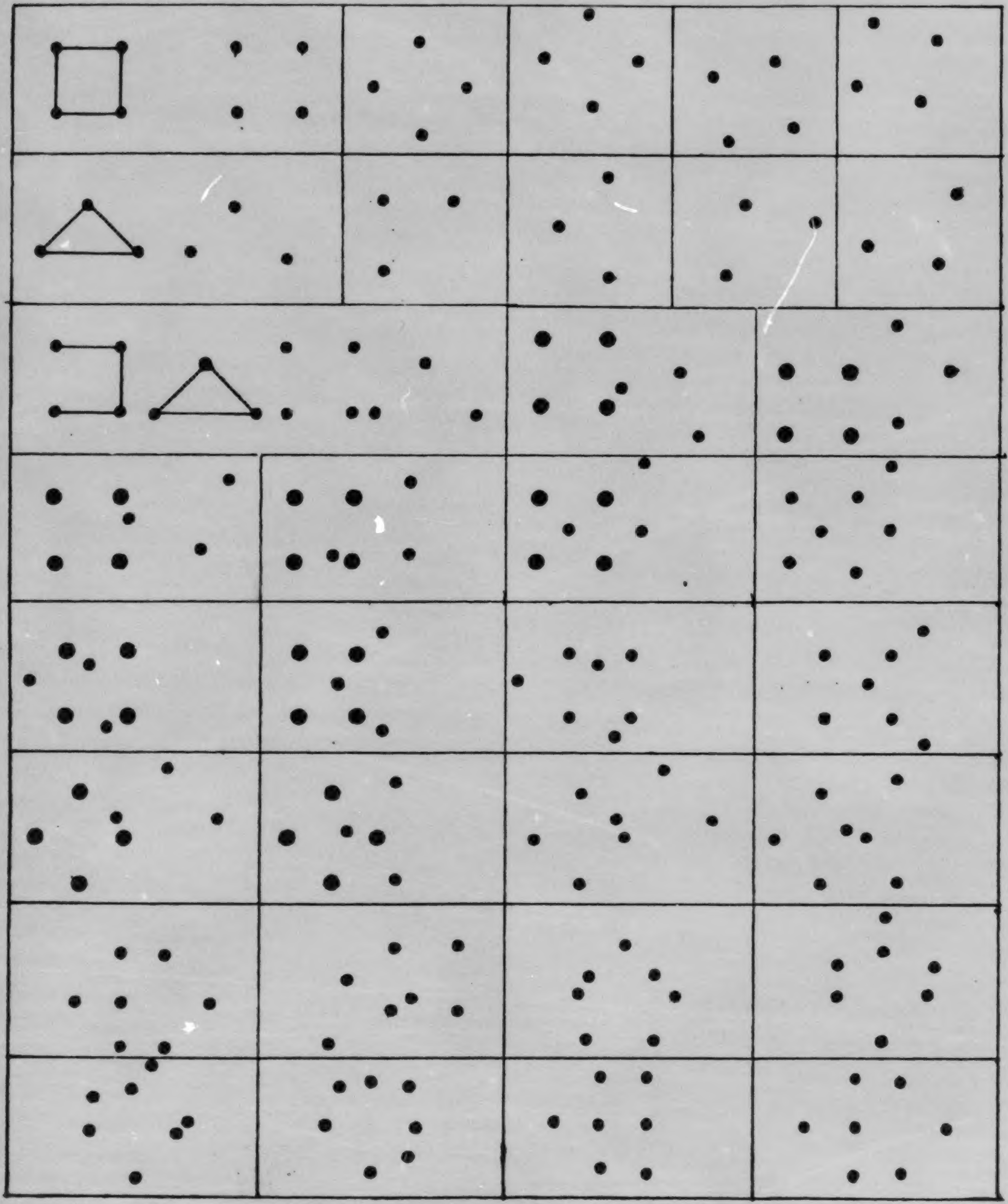




NIE STUDY : Dots Training Sheet, EXERCISE 4

Name: ______________________________ I.D. ____________

School: __________________ Grade: ________ Date: __________________

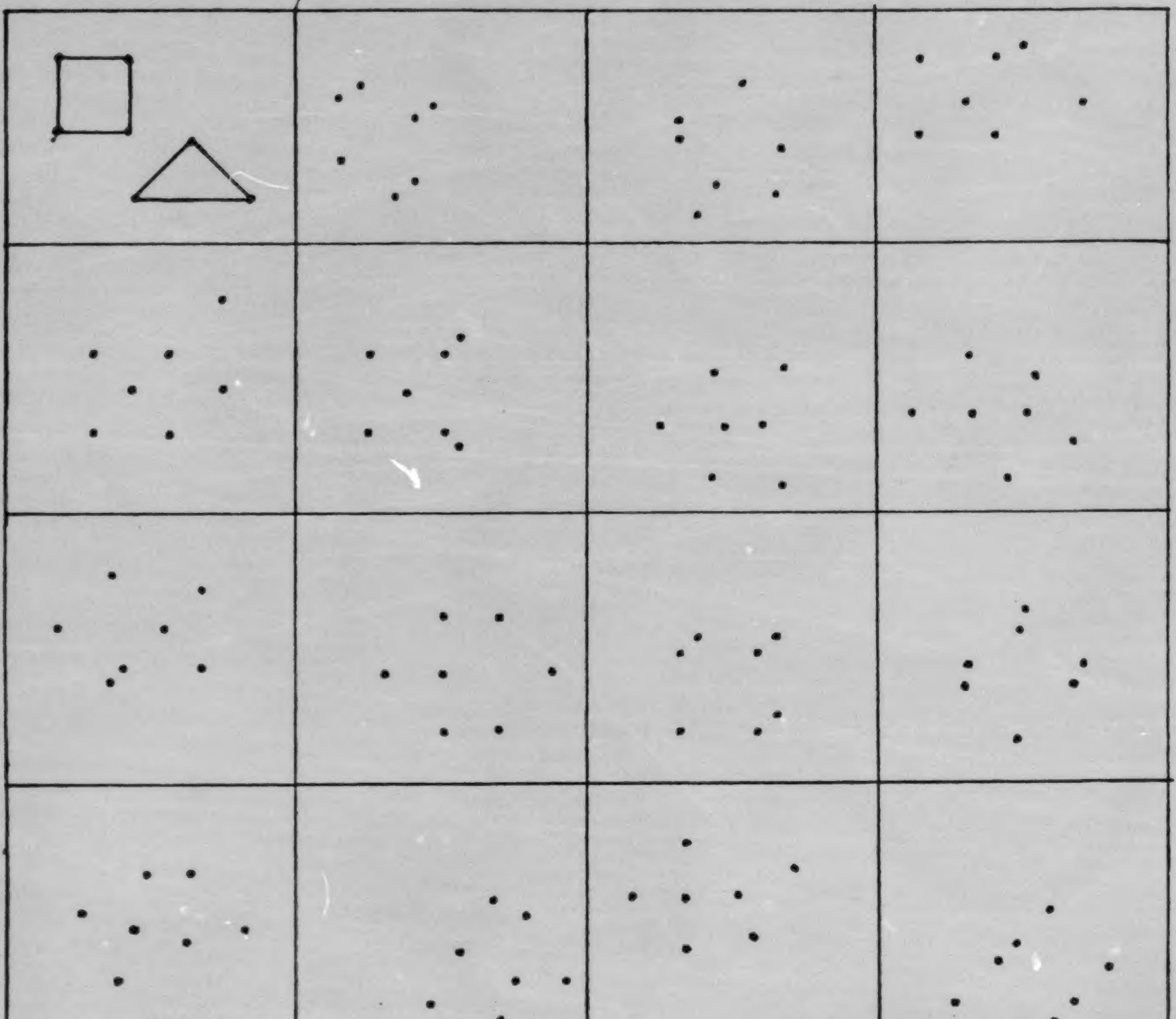


Exercise 4: Dot's Criterion mastery Test

Directions: connect the dots to form one triangle and one square. Each dot belongs to only one shape.

Name ______________________

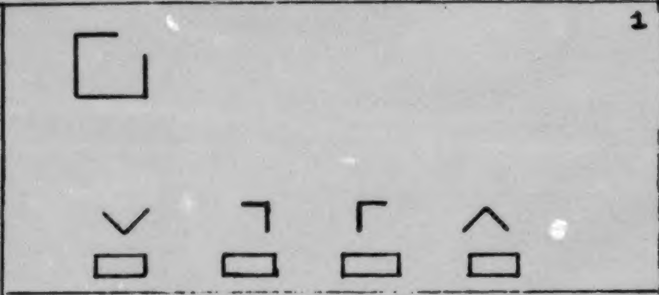
School ______________ GRADE ________ Date ________



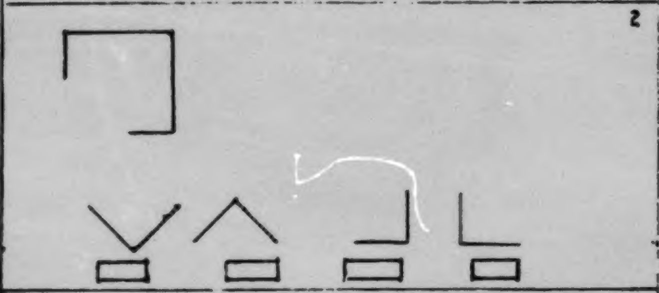
Exercise 5: Figure Completion

____
Name (I.D.)

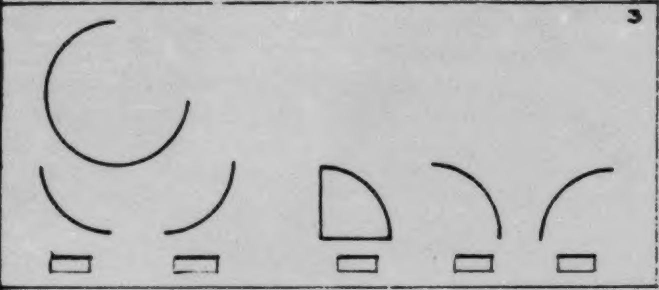
1



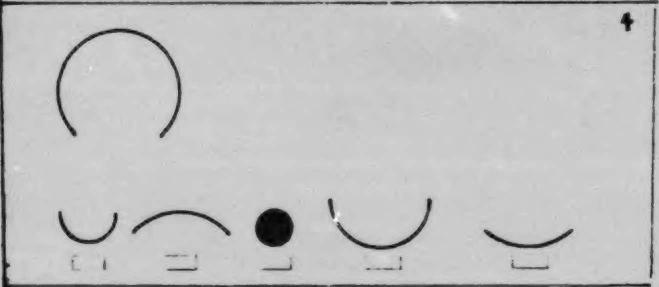
2



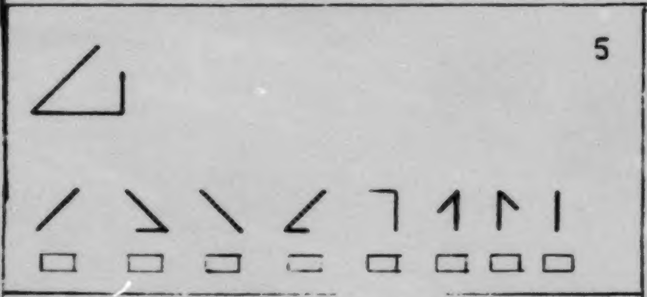
3



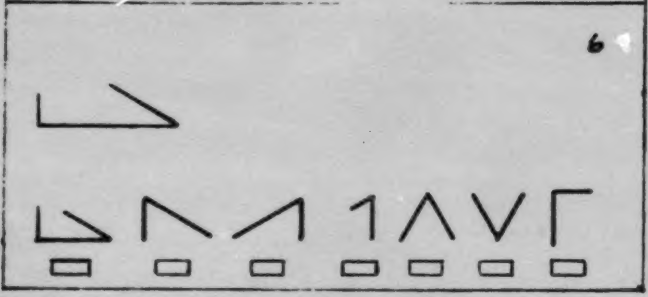
4



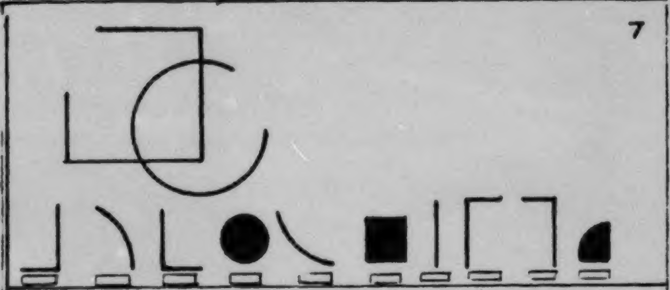
5



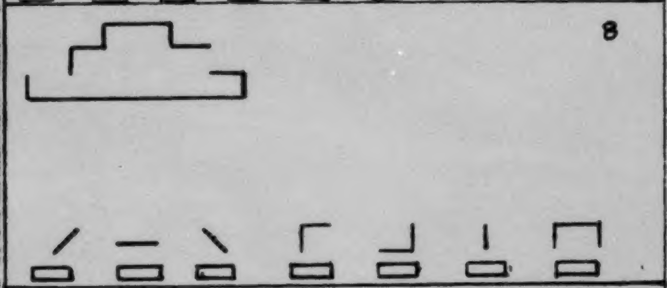
6



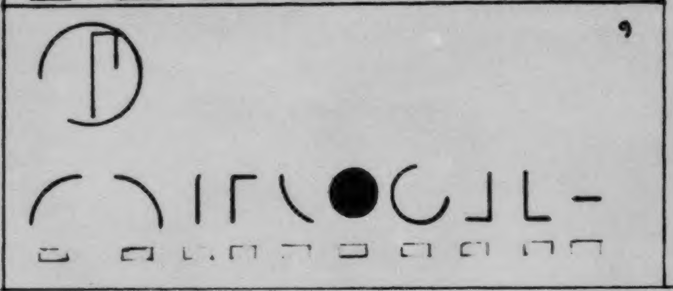
7



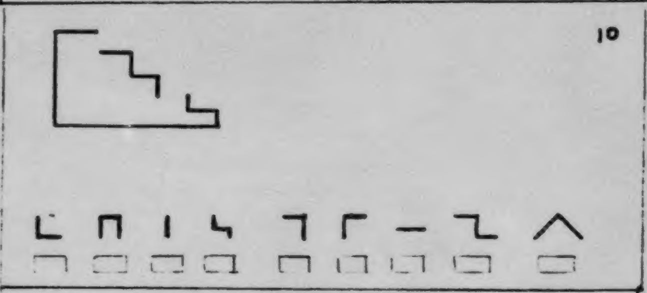
8



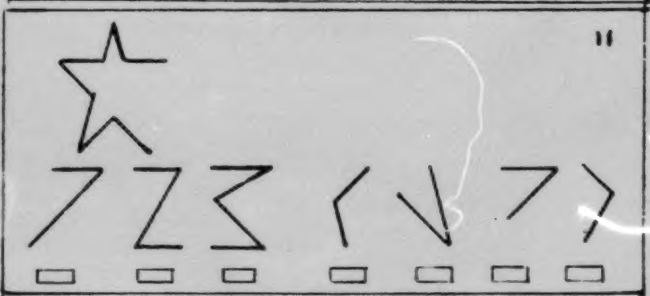
9



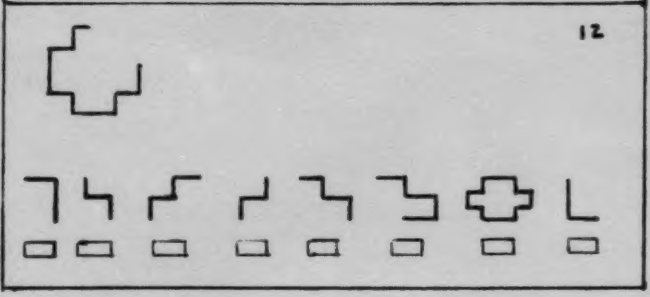
10



11



12

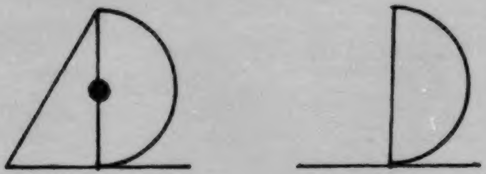


____
School

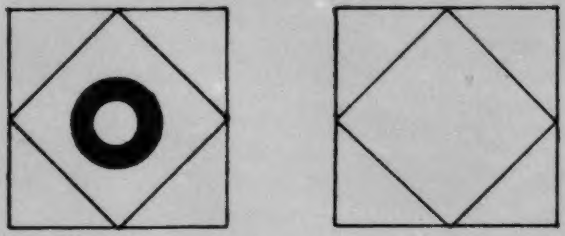
Exercise 6: Pattern Completion

Name ____________________

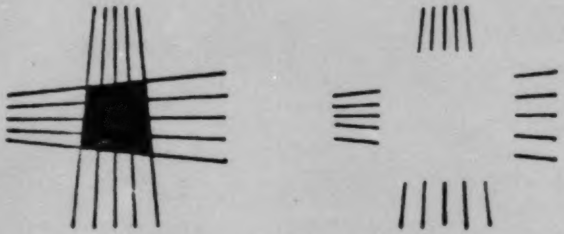
School ____________________



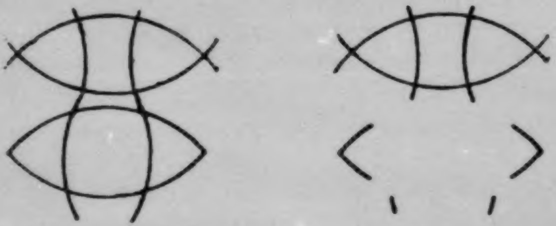
1



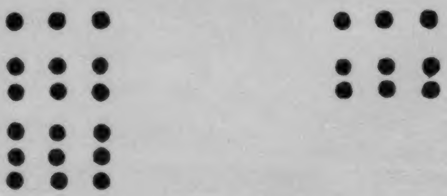
2



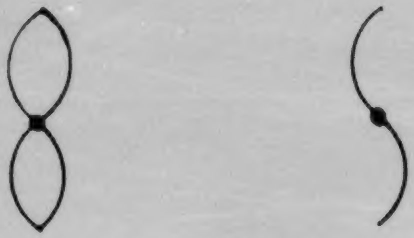
3



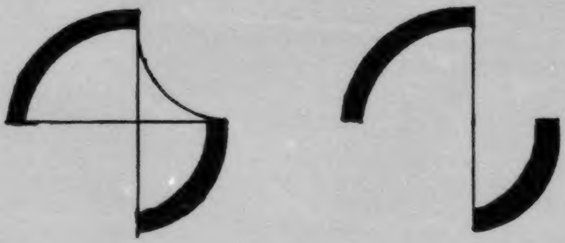
4



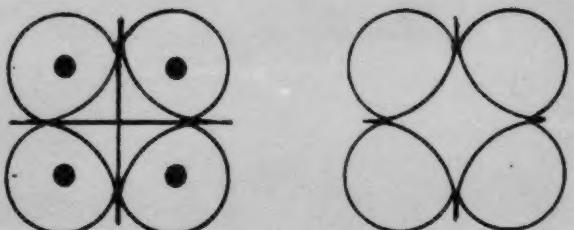
5



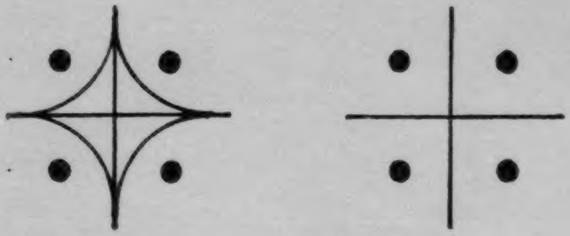
6



7



8

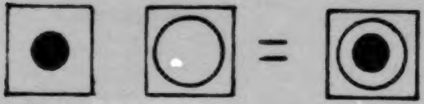


Exercise 7: Combining Patterns

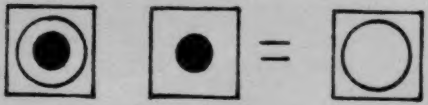
Name

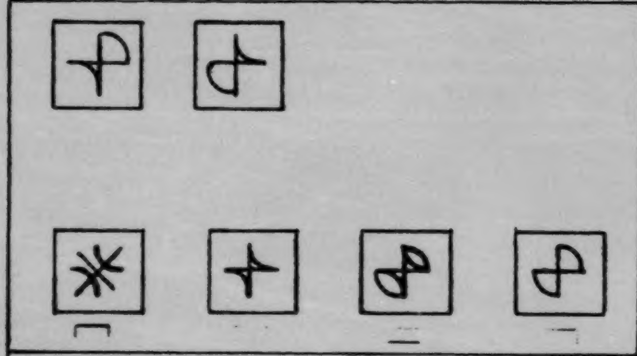
School

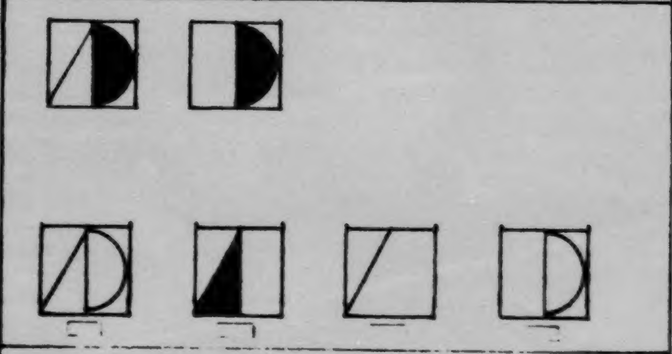
Adding Patterns:

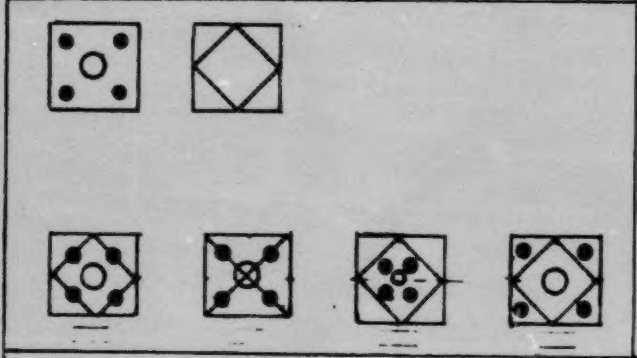


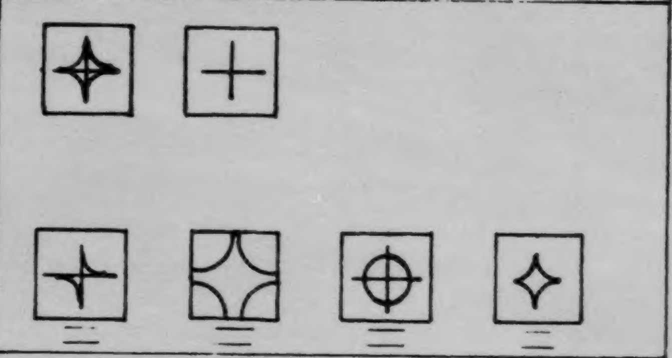
Subtracting Patterns:

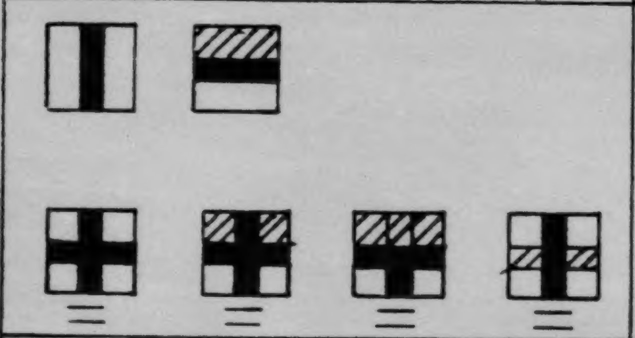


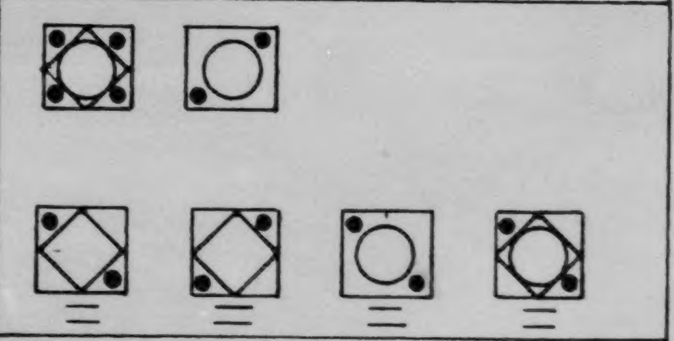


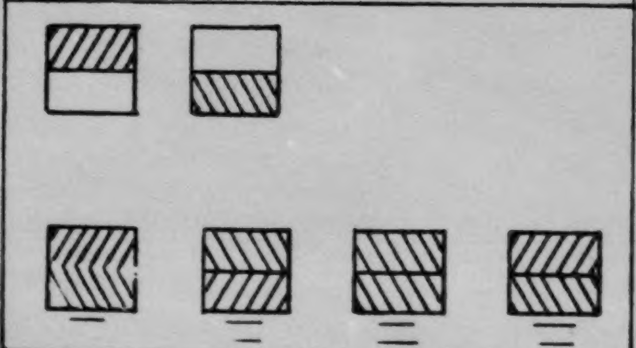


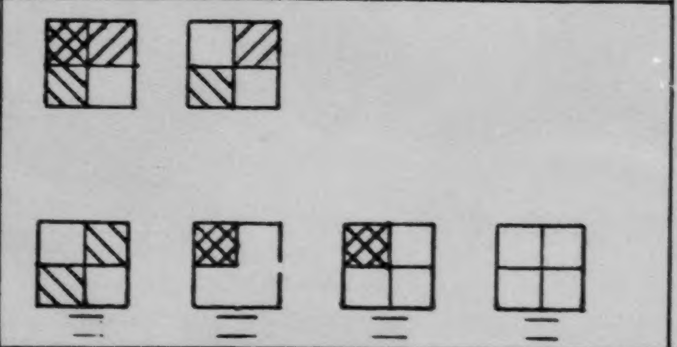










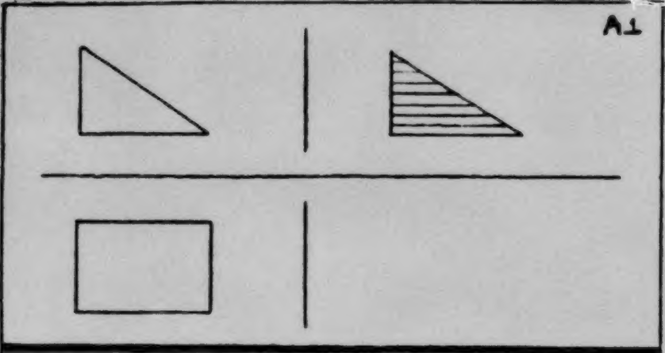


EXERCISE: 8

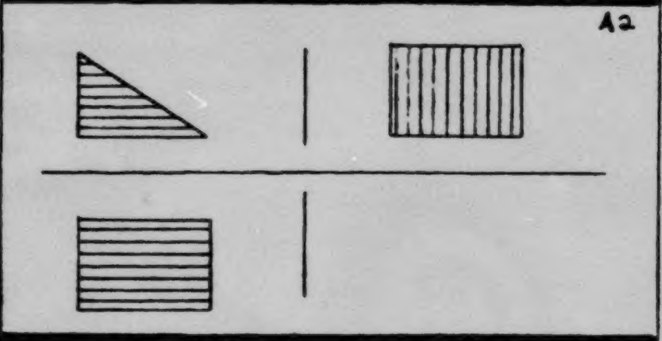
NAME

SCHOOL

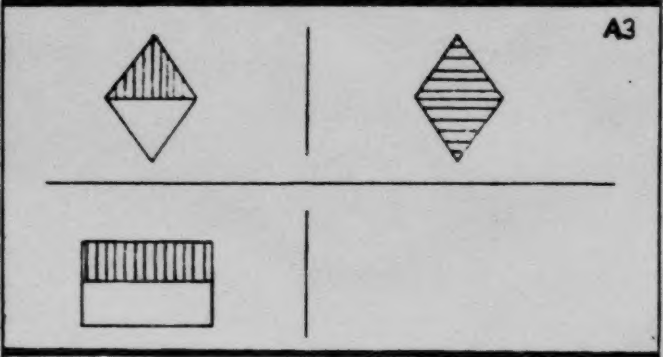
A1



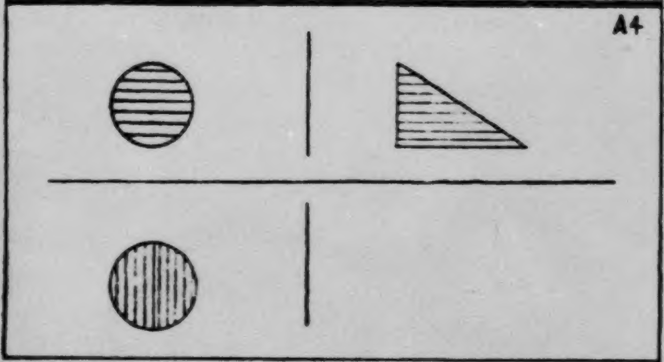
A2



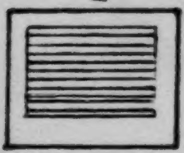
A3



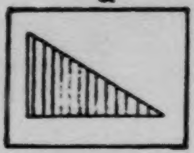
A4



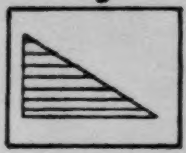
1



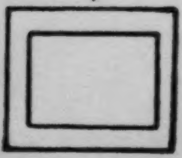
2



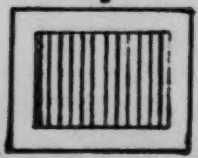
3



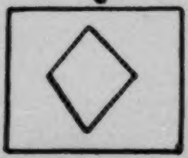
4



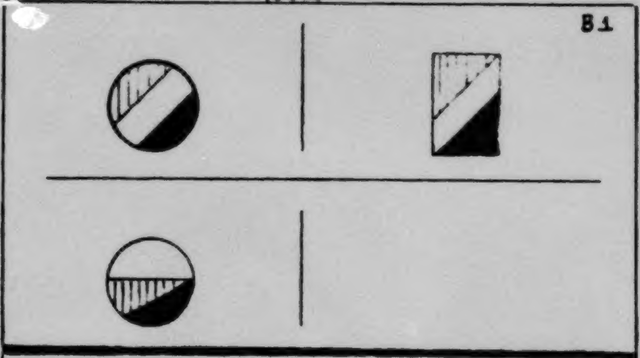
5



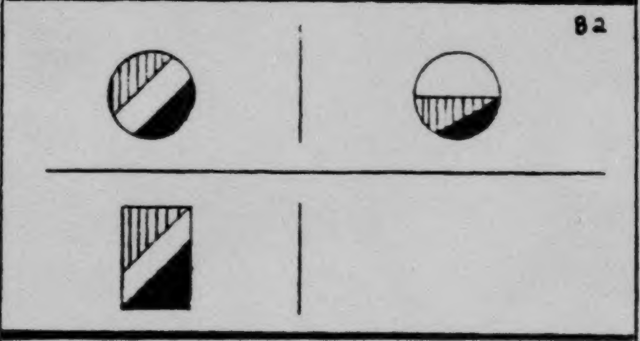
6



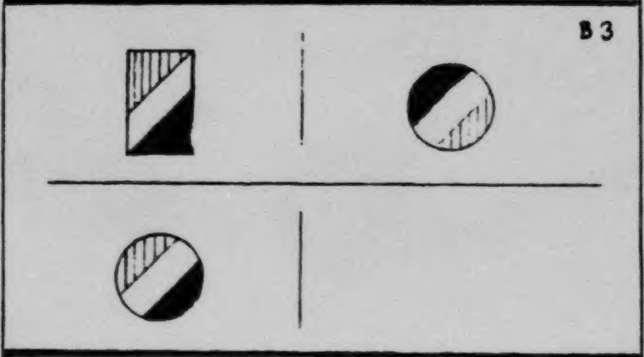
B1



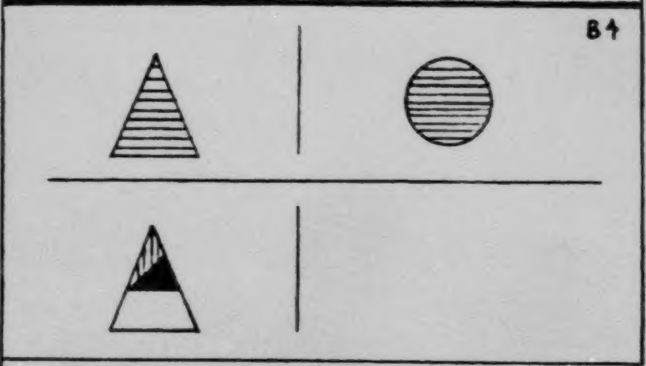
B2



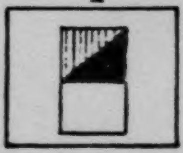
B3



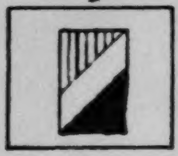
B4



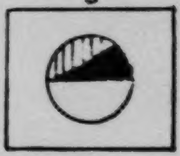
1



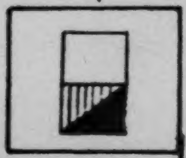
2



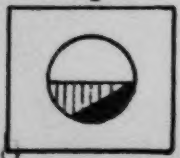
3



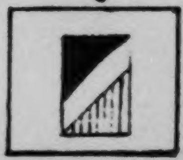
4



5



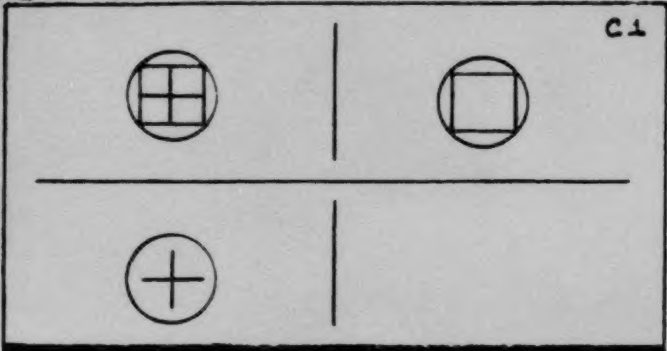
6

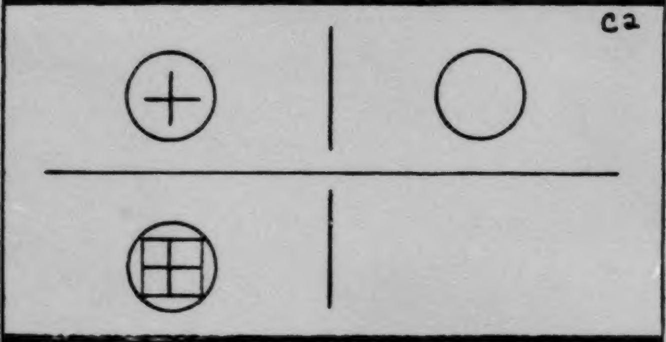


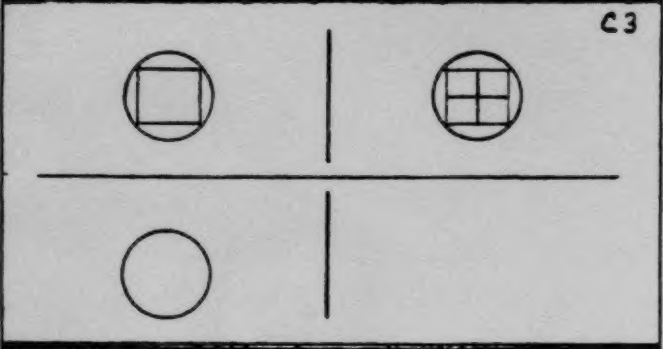
EXERCISE: 8

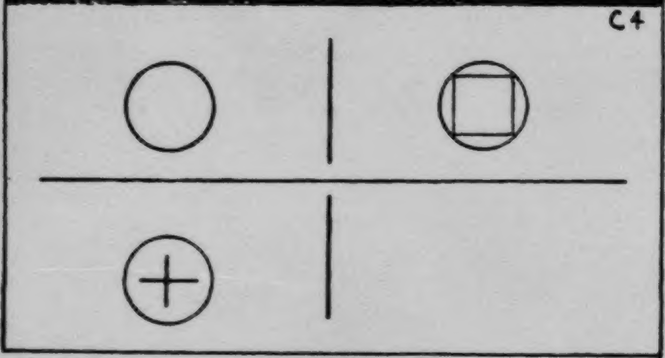
NAME

SCHOOL

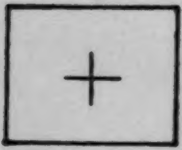




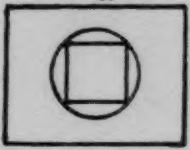




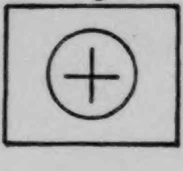
1



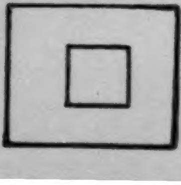
2



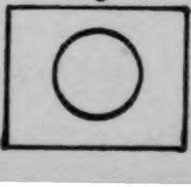
3



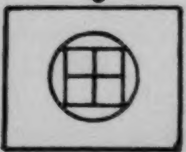
4



5



6



1

2

3

4

1

2

3

4

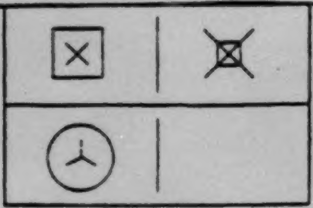
5

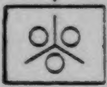
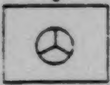
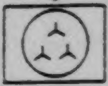
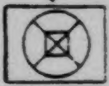
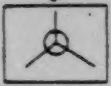
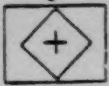
6

Exercise 9: Analogies Criterion Test

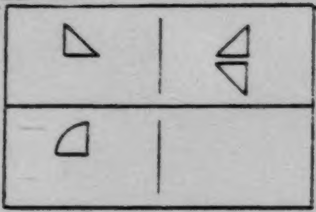
Name

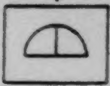
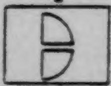
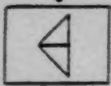
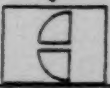
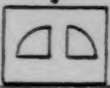
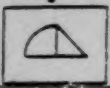
1



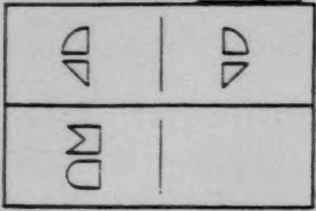
1

2

3

4

5

6


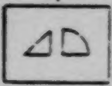
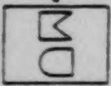
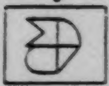
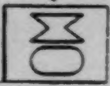
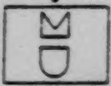
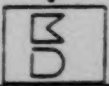
2



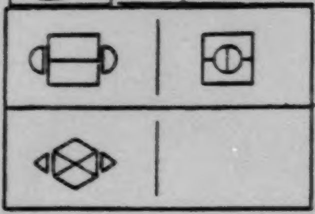
1

2

3

4

5

6


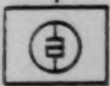
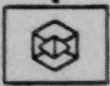
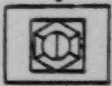
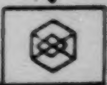
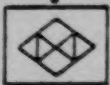
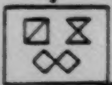
3



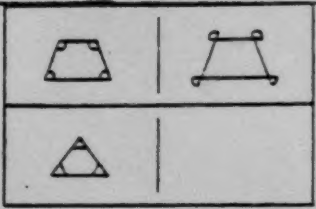
1

2

3

4

5

6


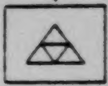
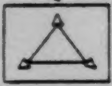
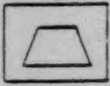
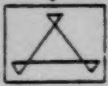
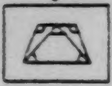
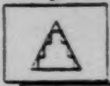
4



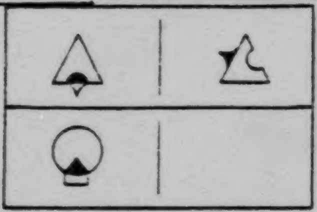
1

2

3

4

5

6


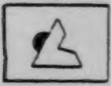
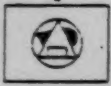
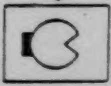
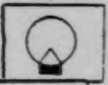
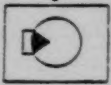
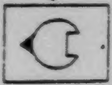
5



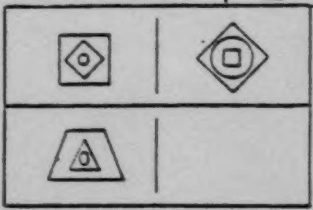
1

2

3

4

5

6


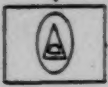
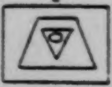
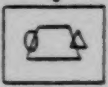
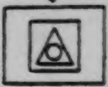
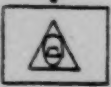
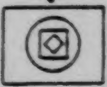
6



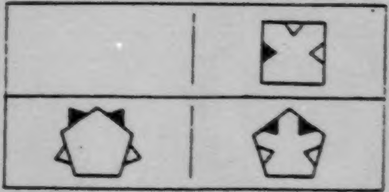
1

2

3

4

5

6


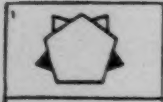
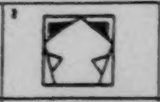
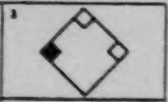
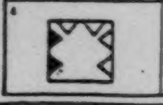
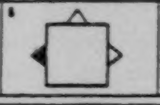
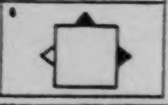
7



1

2

3

4

5

6


8



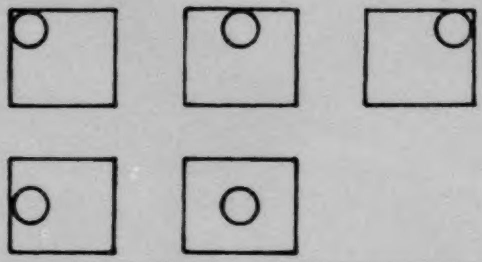
1

2

3

4

5

6


School

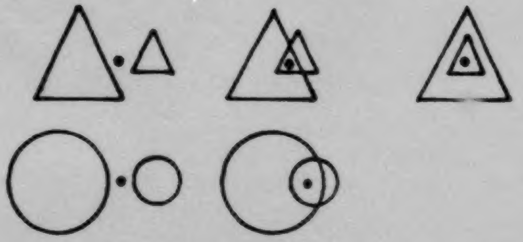
EXERCISE 10: 2×3 ANALOGIES

NAME

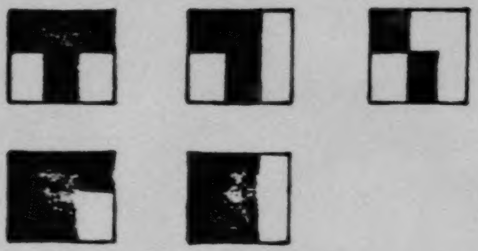
1



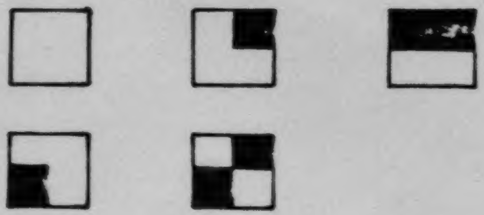
2



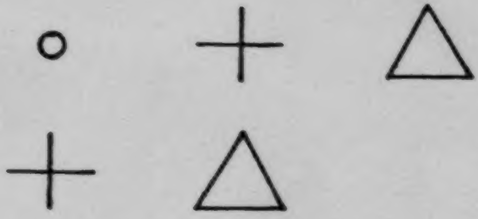
3



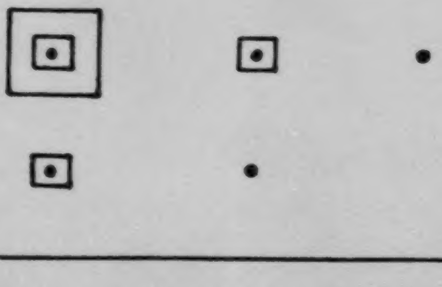
4



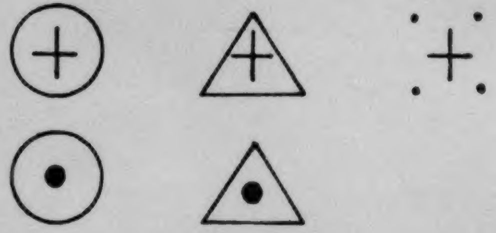
5



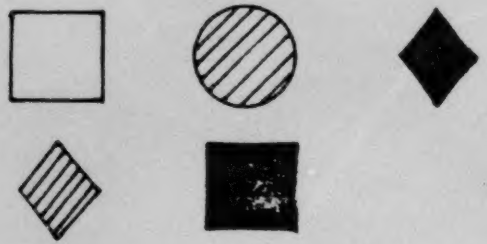
6



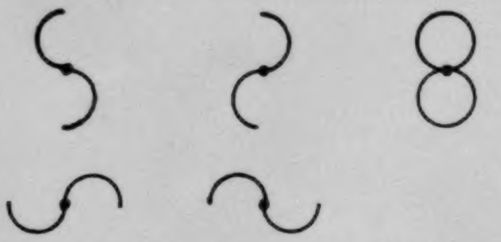
7



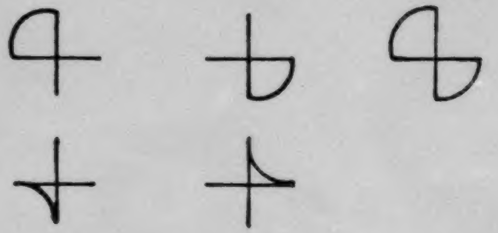
8



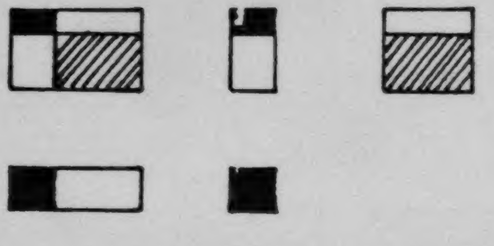
9



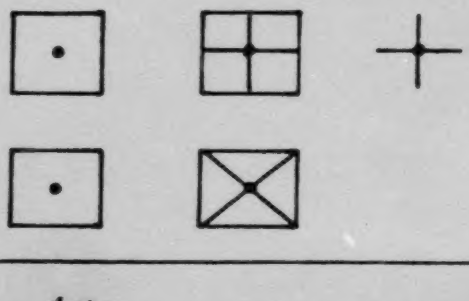
10



11



12



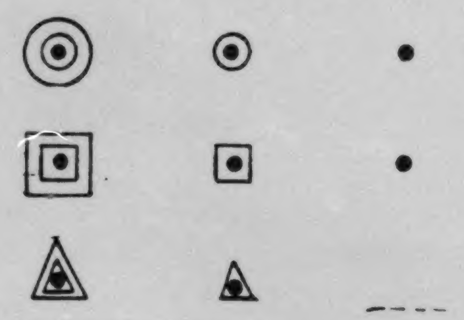
School

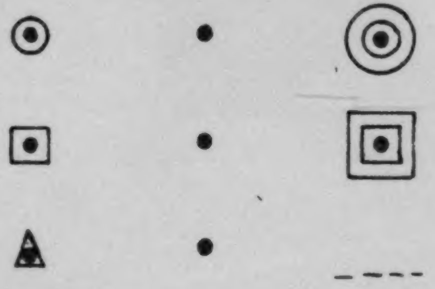
EXERCISE: III

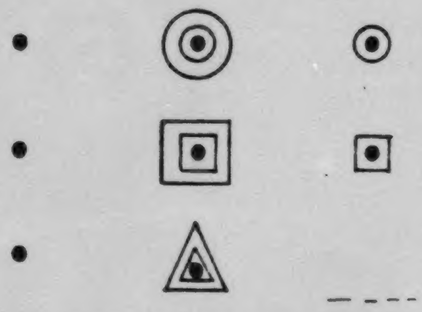
P:

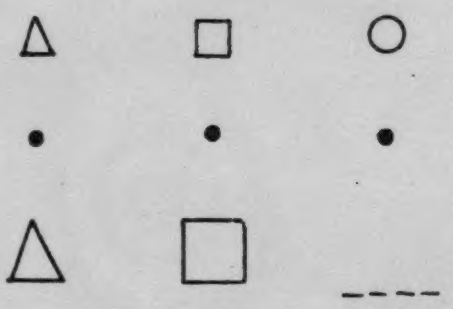
AME

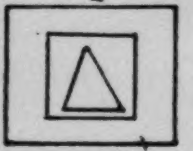
SCHOOL

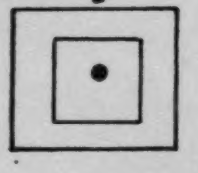
1


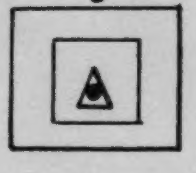
2


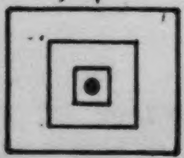
3


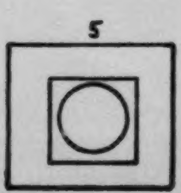
4


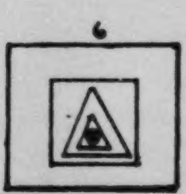
1


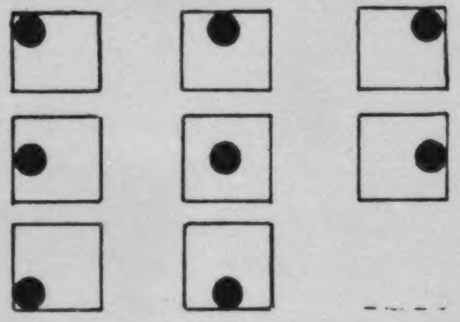
2


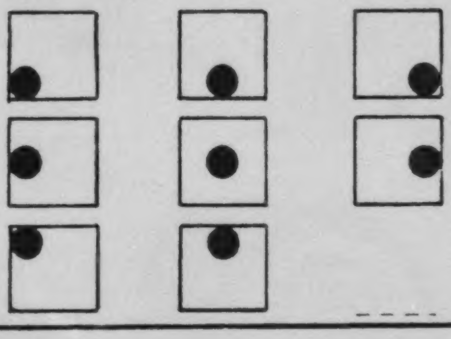
3


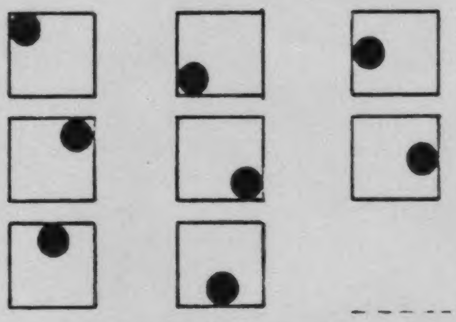
4


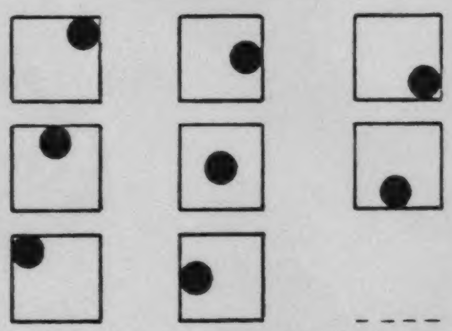
5


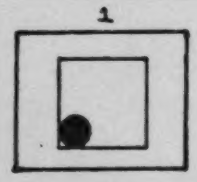
6


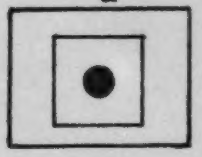
1


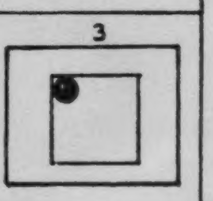
2


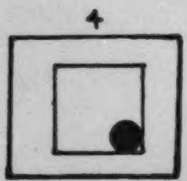
3


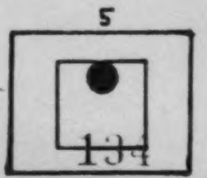
4


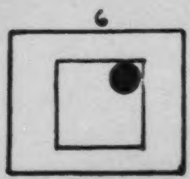
1


2


3


4


5


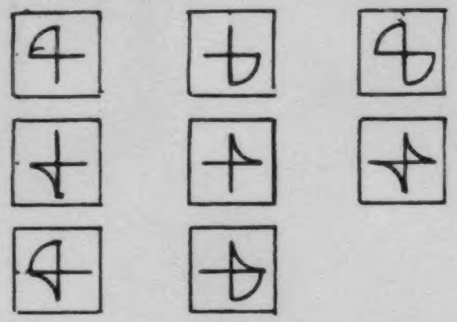
6


EXERCISE: 11

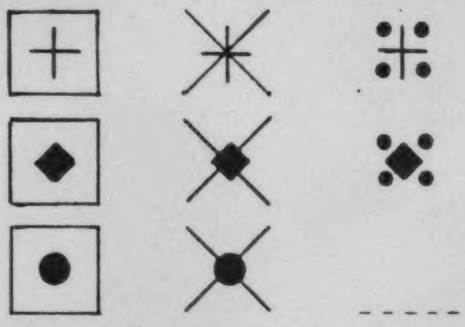
AME

SCHOOL

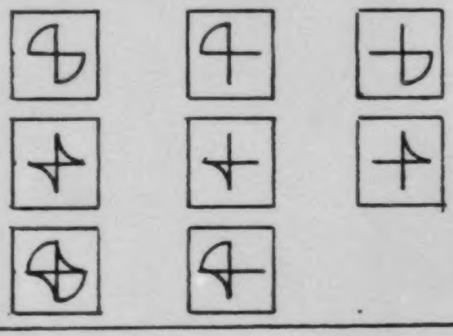
1



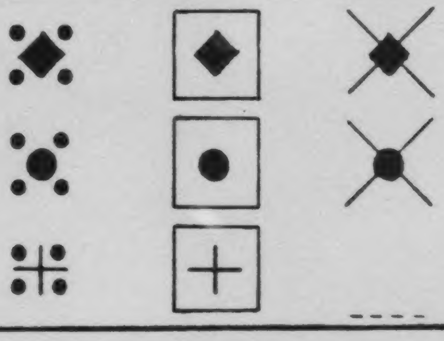
1



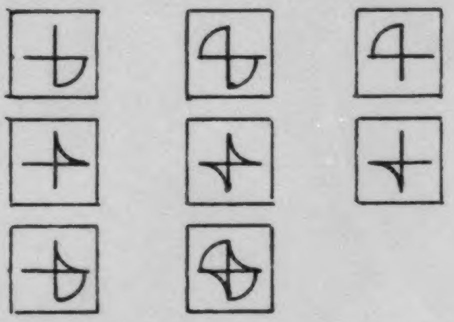
2



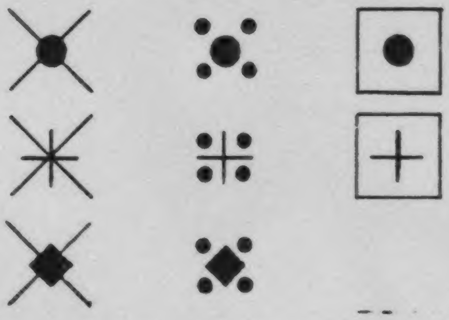
2



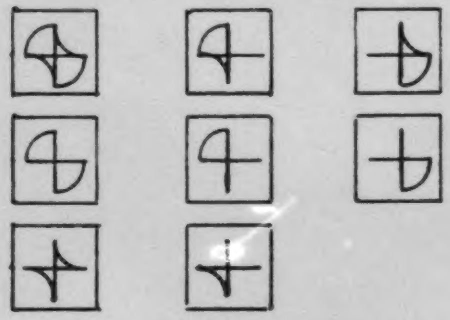
3



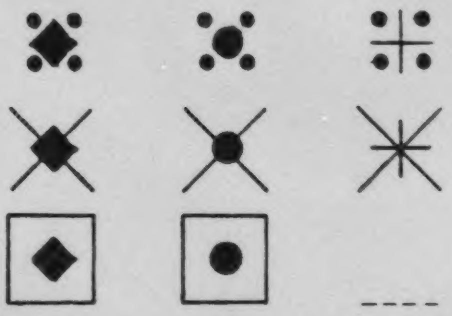
3



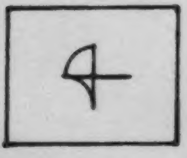
4



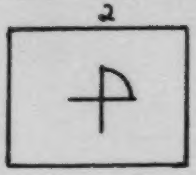
4



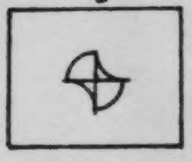
1



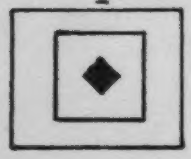
2



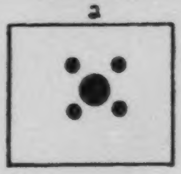
3



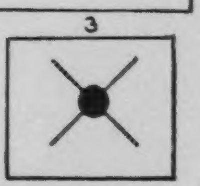
1



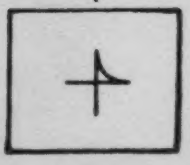
2



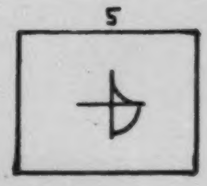
3



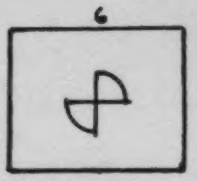
4



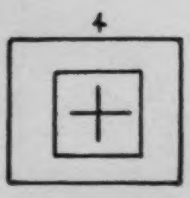
5



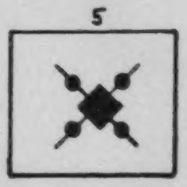
6



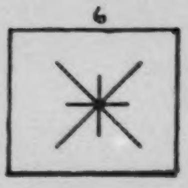
4



5



6

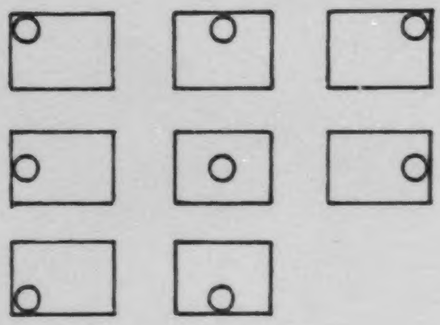


EXERCISE: 12 3x3 CRITERION MASTERY

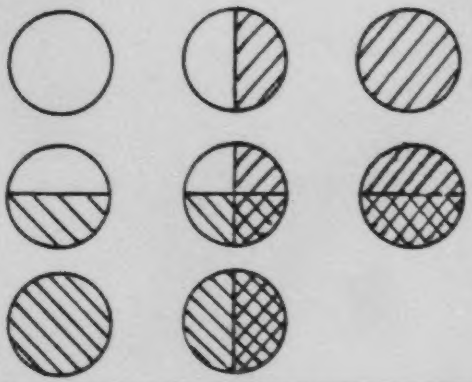
SCHOOL ____________

NAME ____________

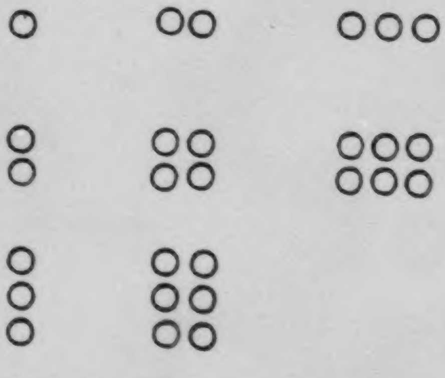
1



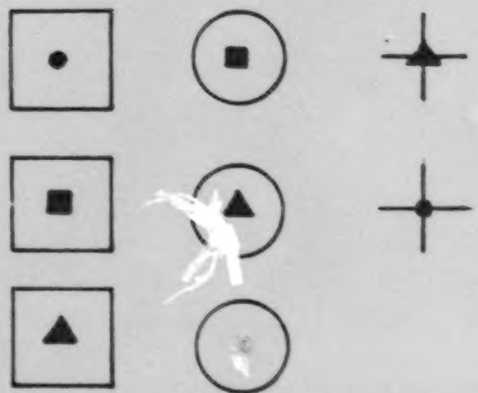
2



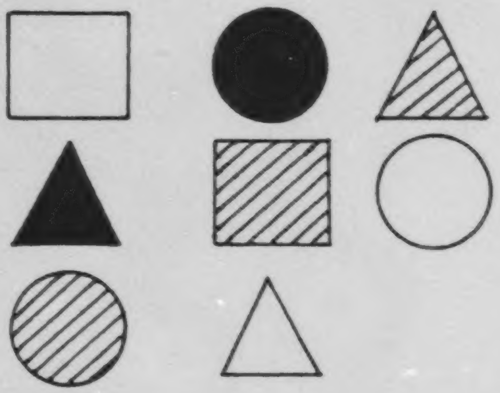
3



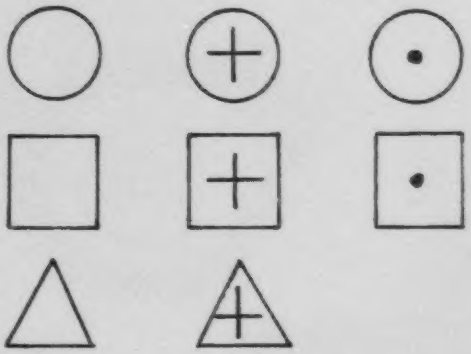
4



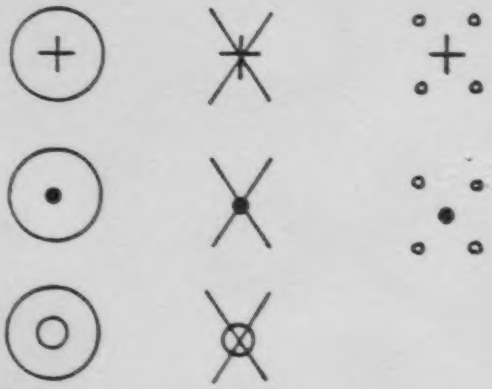
5



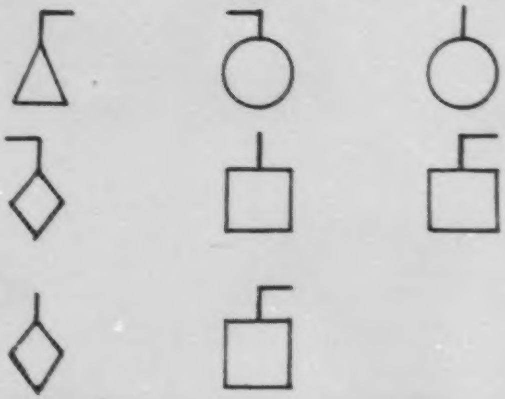
6



7



8

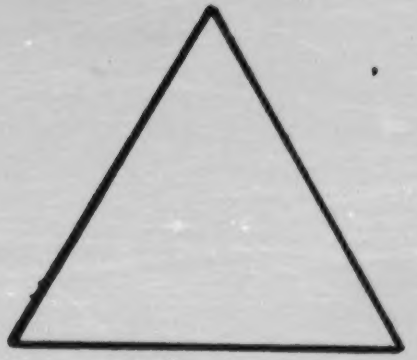


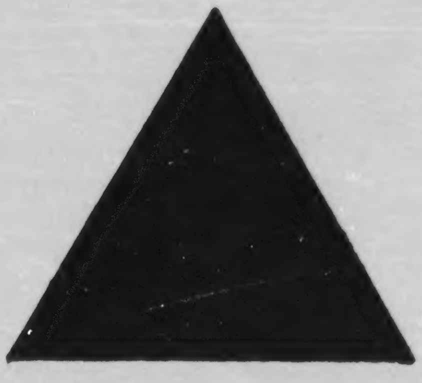
APPENDIX B

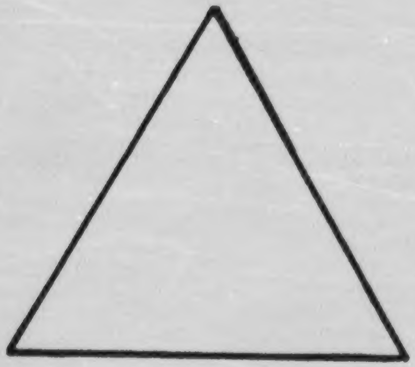
MEDIATED LEARNING EXERCISES

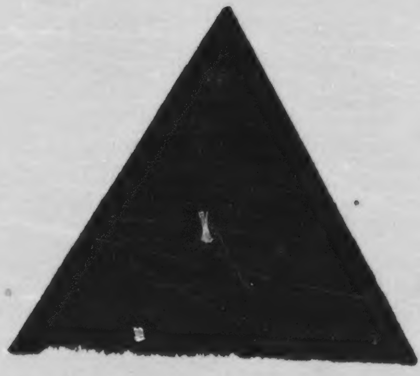
EXERCISE 1

1

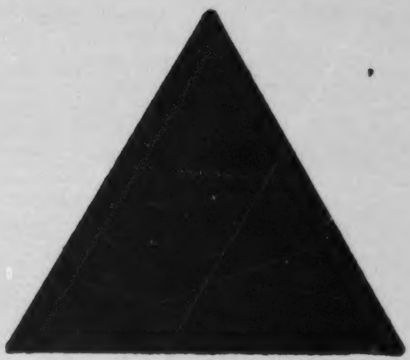

COLOR


☐

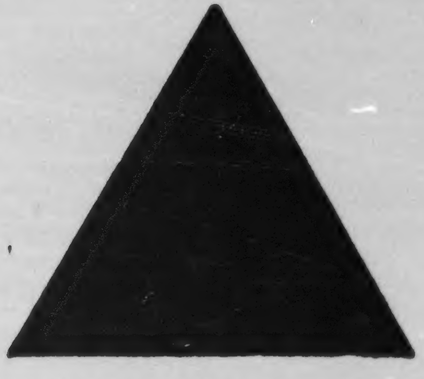

☐


☐

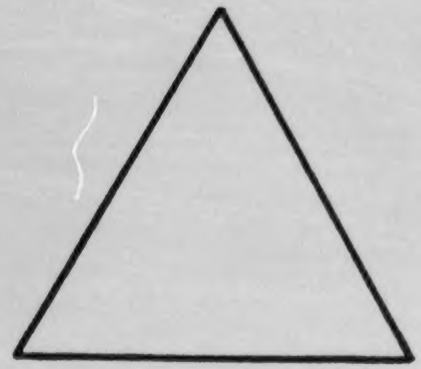
2



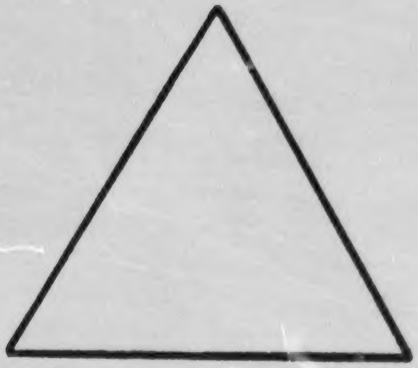
COLOR



☐



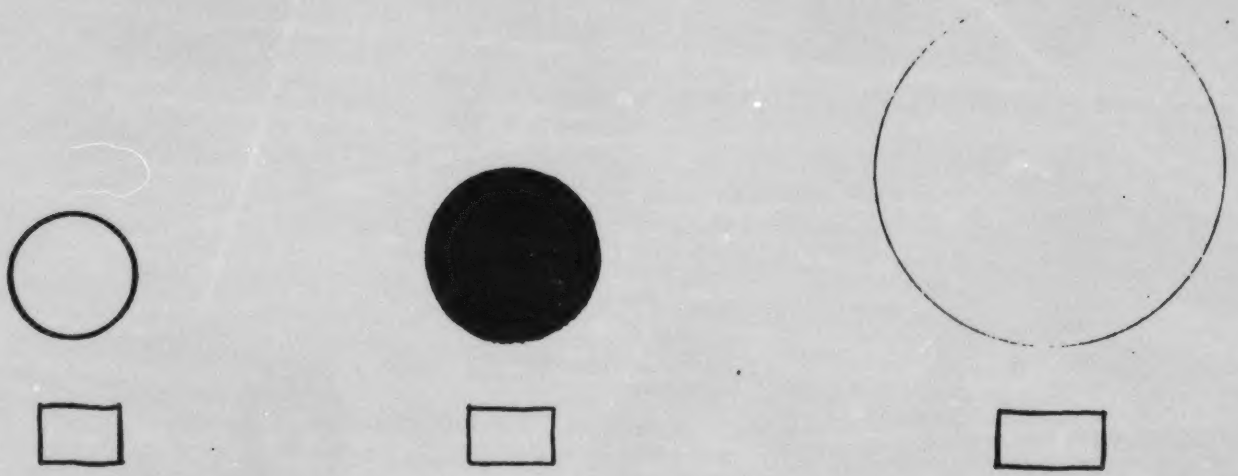
☐



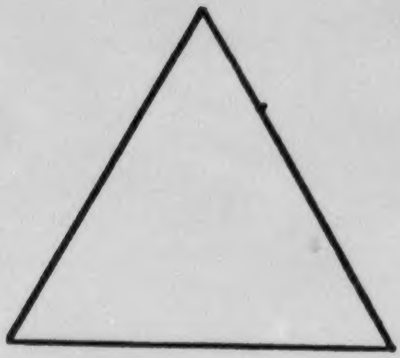
☐

COLOR

142

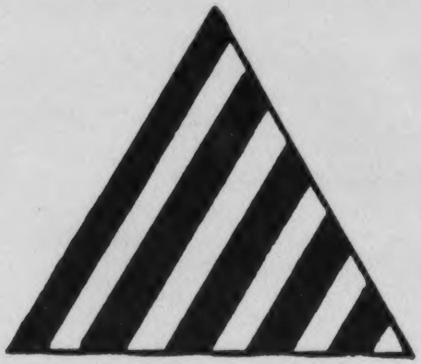


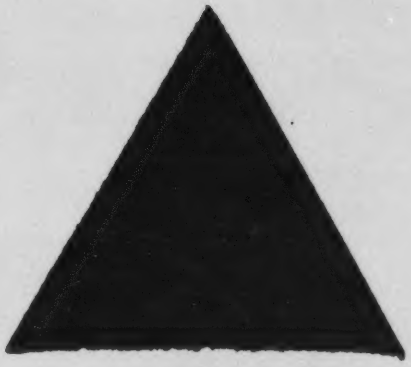
143

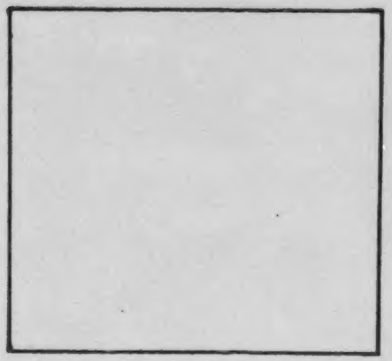


COLOR

144

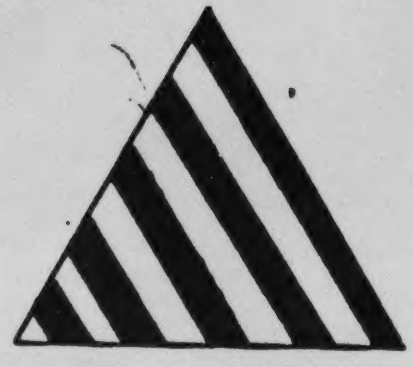




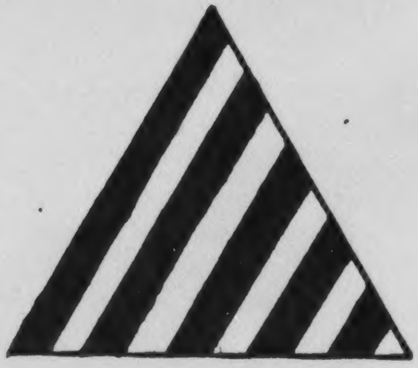


☐ ☐ ☐

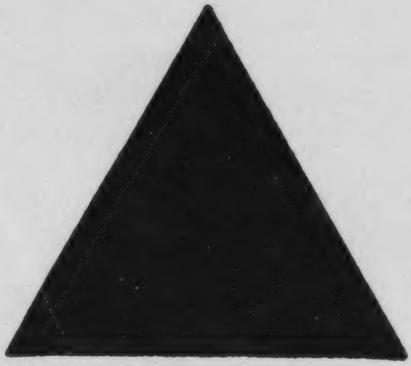
145



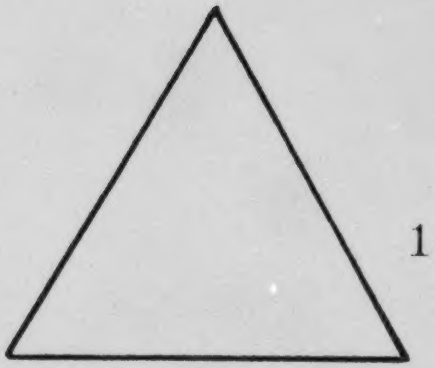
COLOR



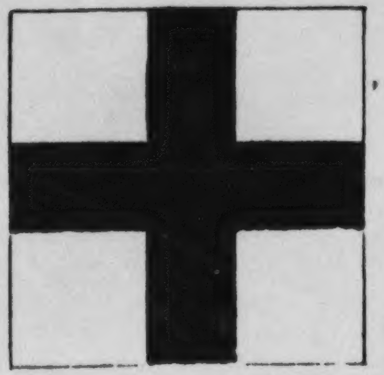
☐



☐

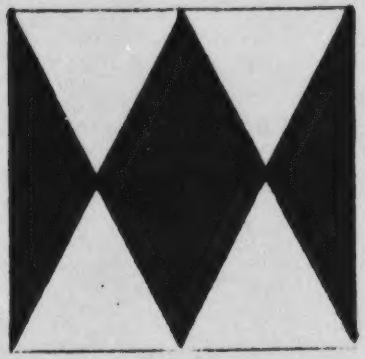


☐

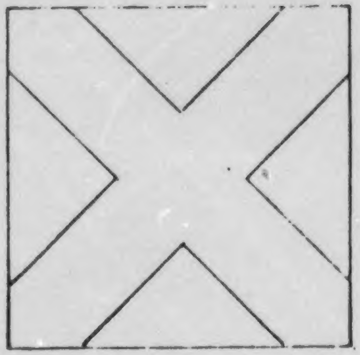


COLOR

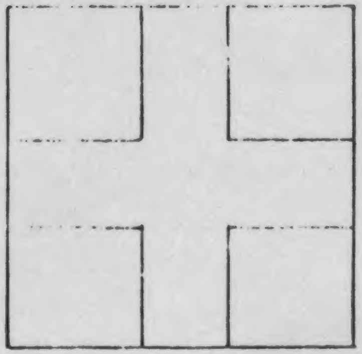
148



☐

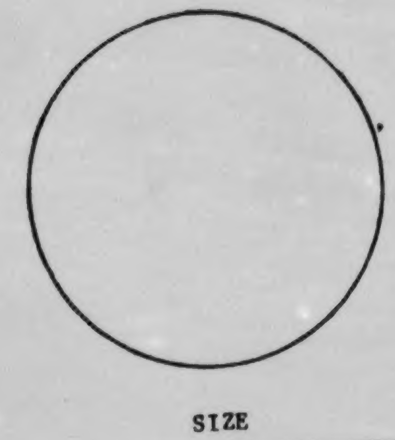


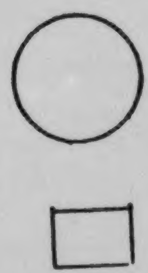
☐

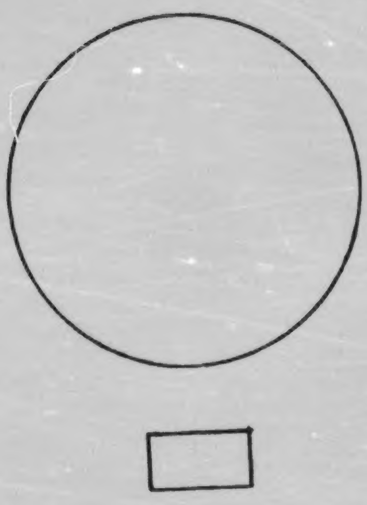


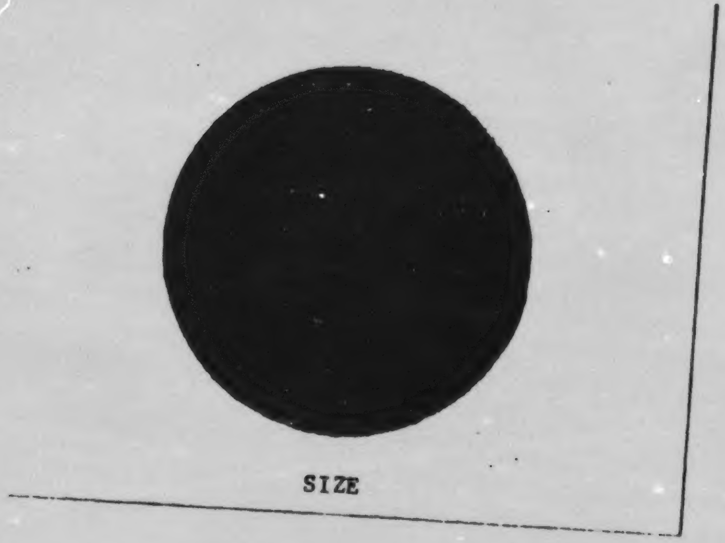
☐

149

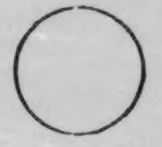

SIZE



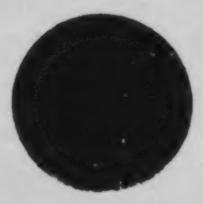




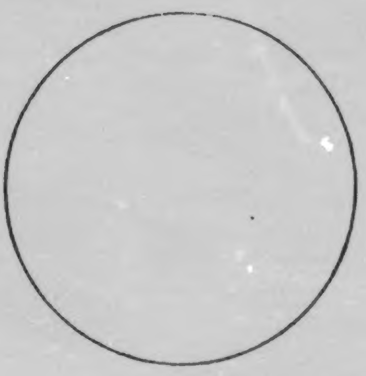
152



☐



☐

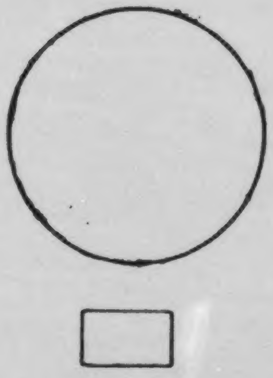


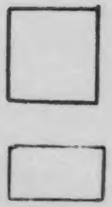
☐

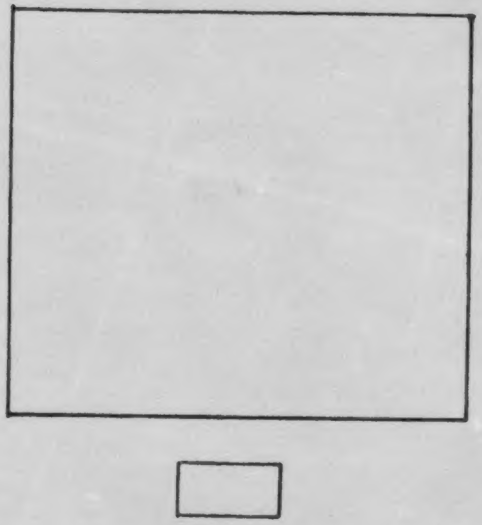
153

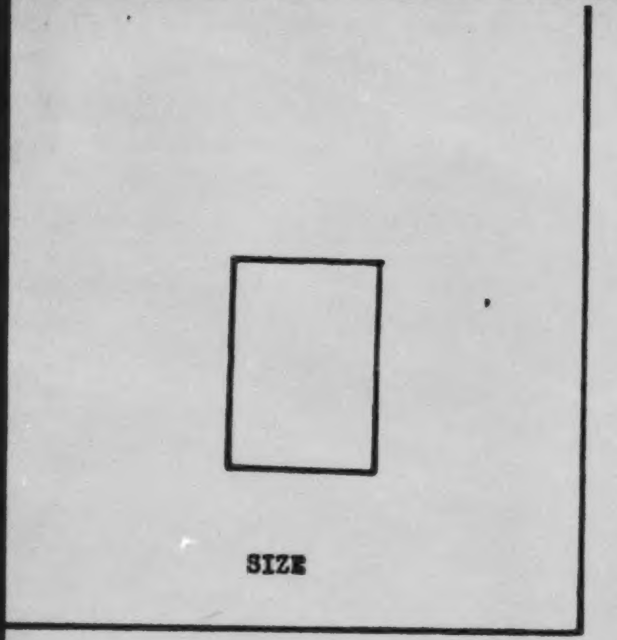
9

SIZE

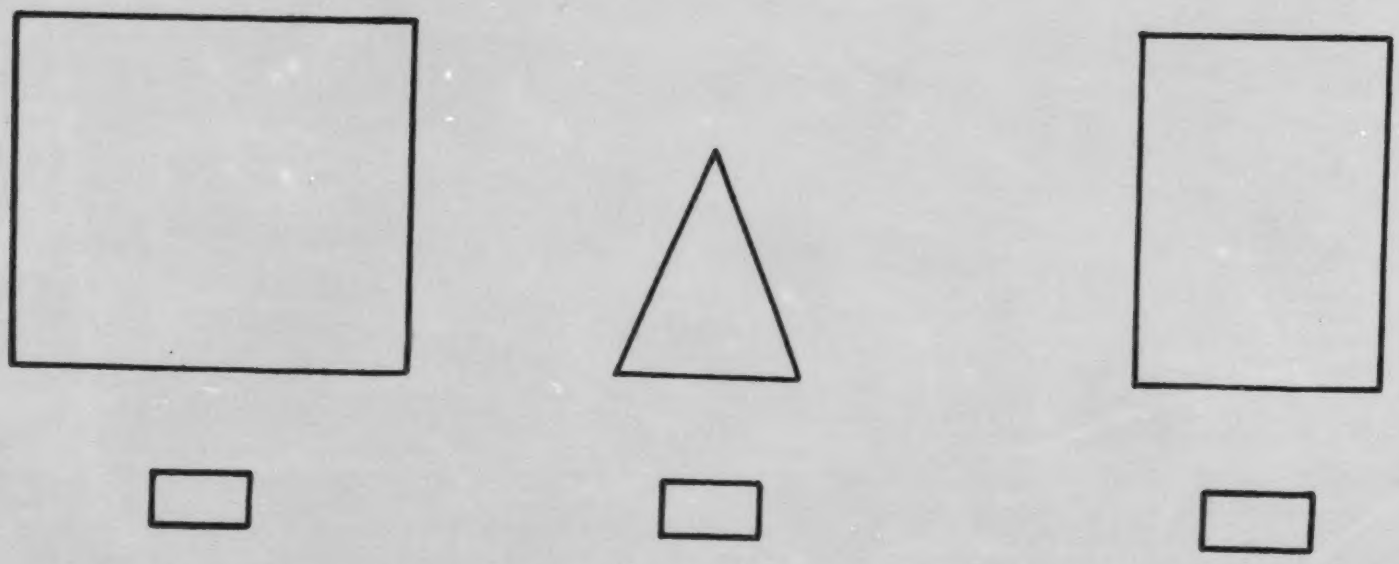




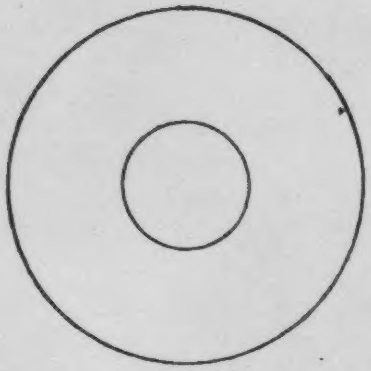




156

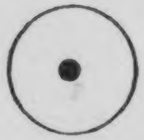


157

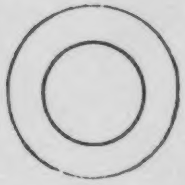


SIZE

158



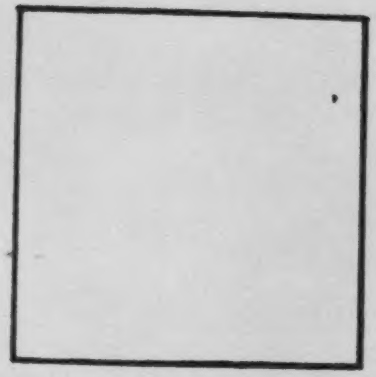
☐



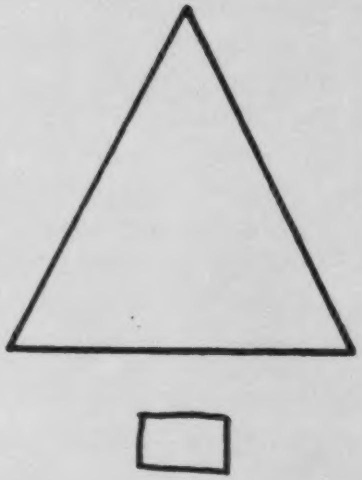
☐

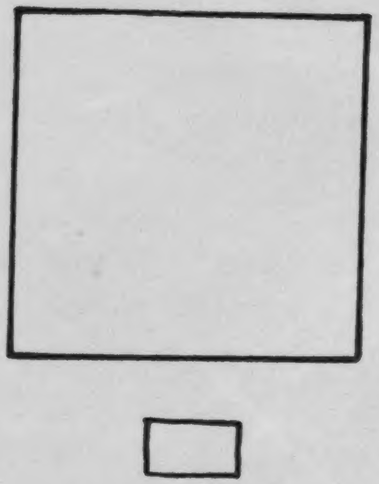
☐

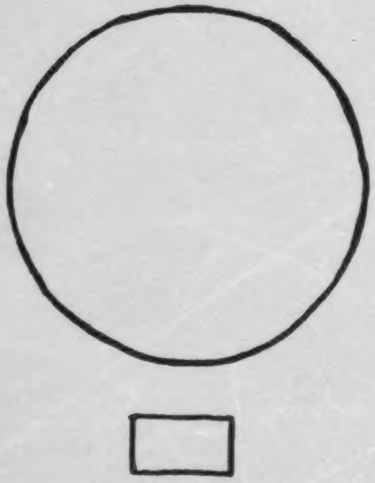
159

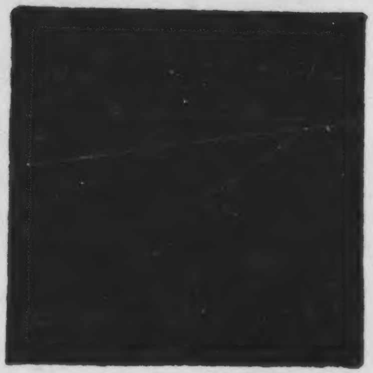


SHAPE



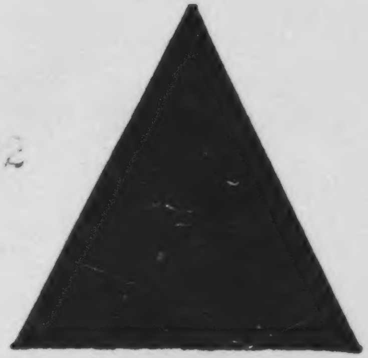




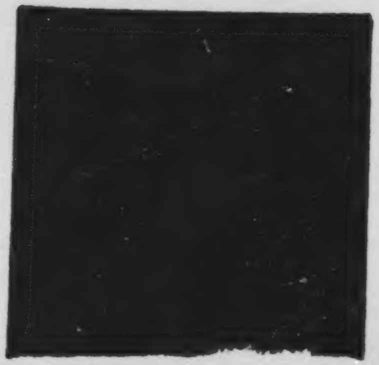


SHAPE

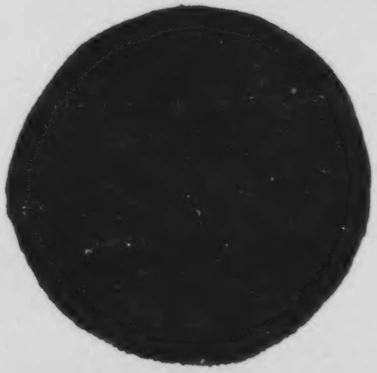
162



☐

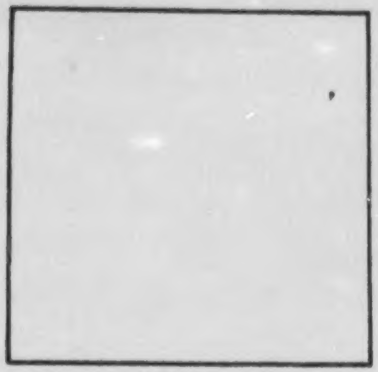


☐



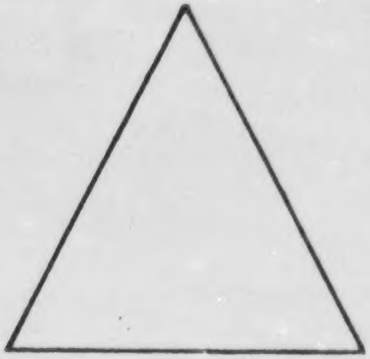
☐

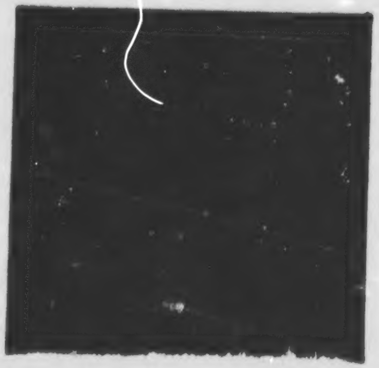
163

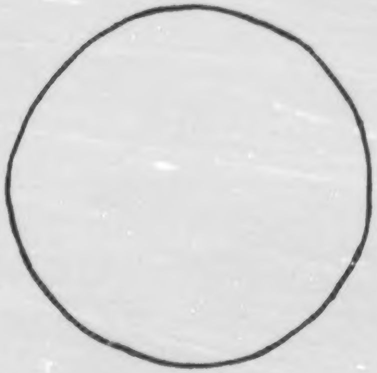


SHAPE

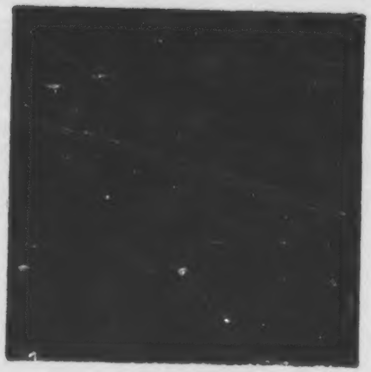
164





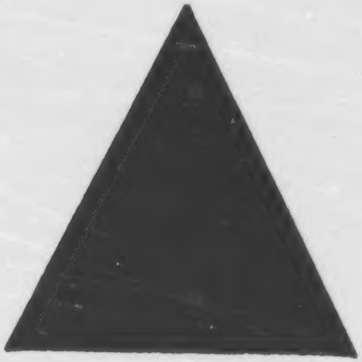


165

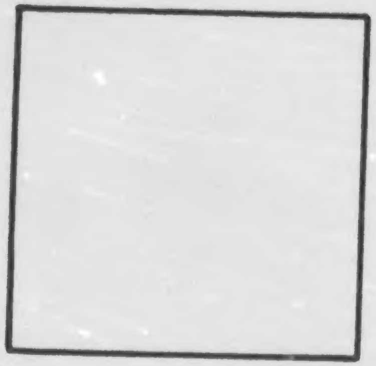


SHAPE

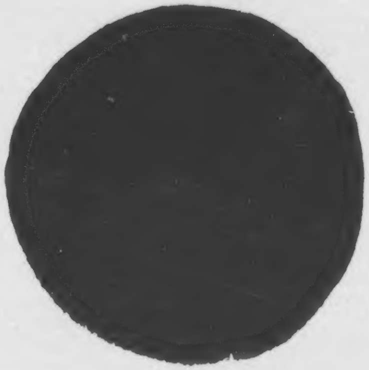
166



☐

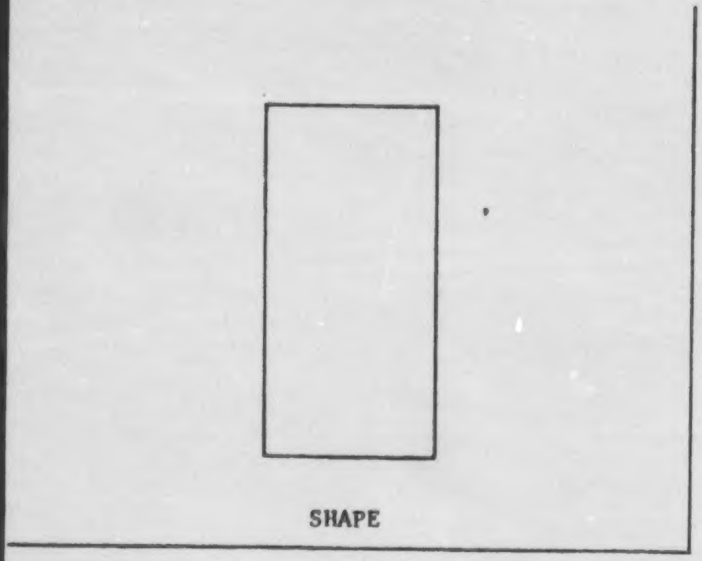


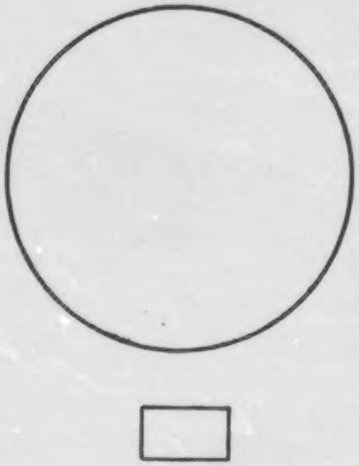
☐

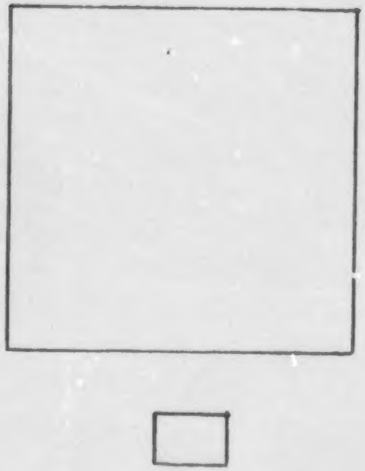


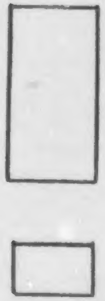
☐

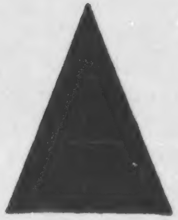
167


SHAPE



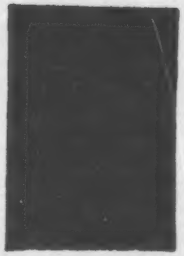




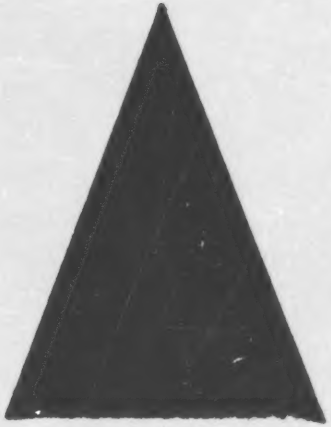


SHAPE

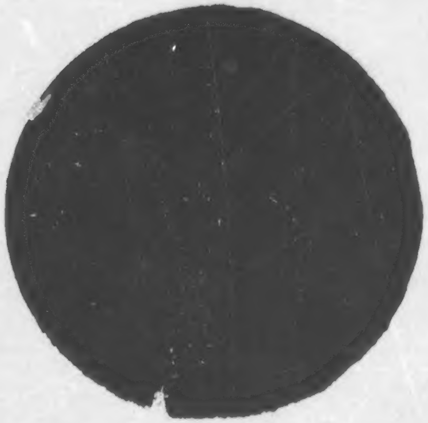
170



☐



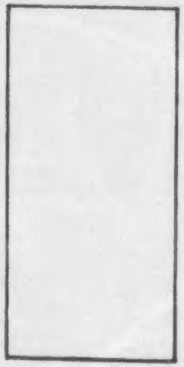
☐



☐

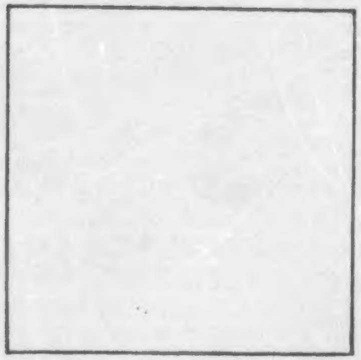
171

18

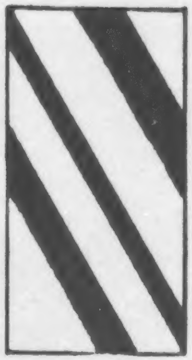


SHAPE

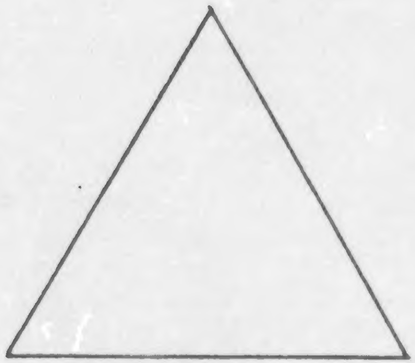
172



☐



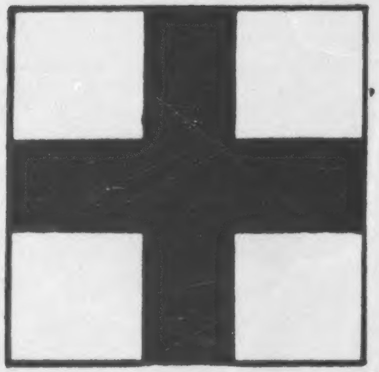
☐



☐

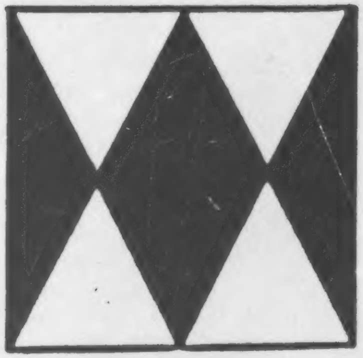
173

19

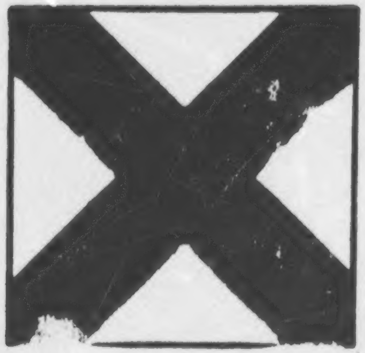


PATTERN

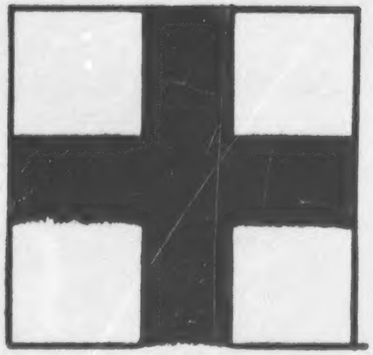
174



☐

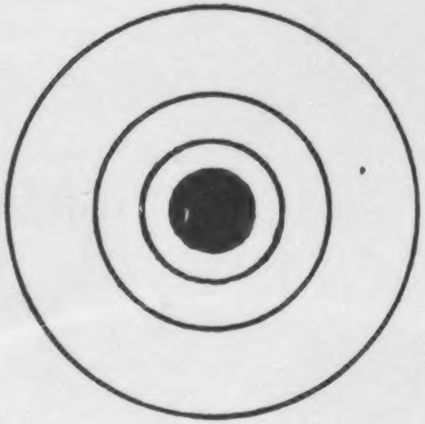


☐



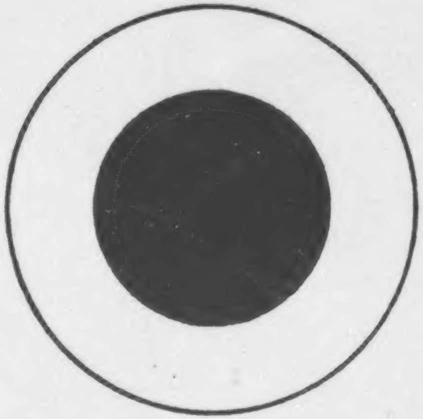
☐

175

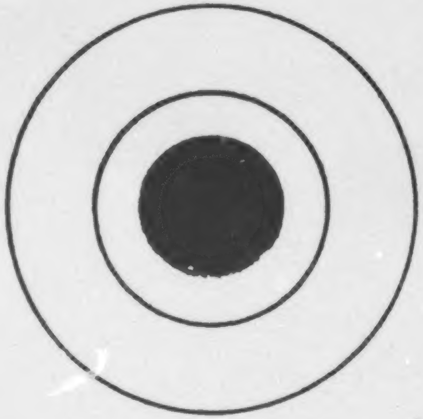


PATTERN

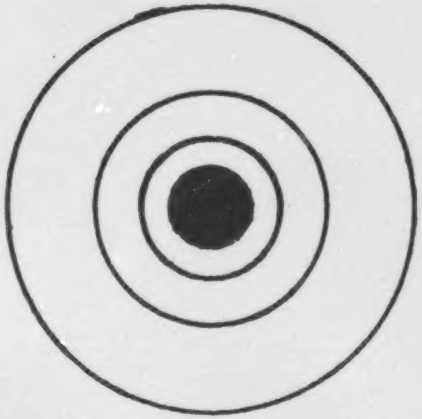
176



☐

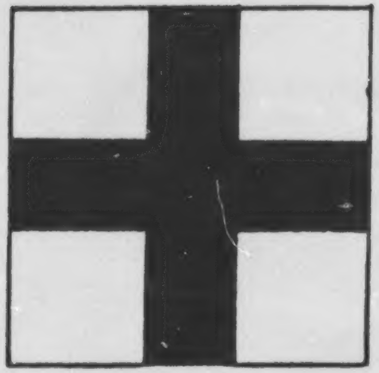


☐



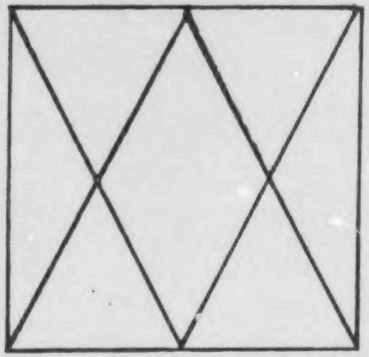
☐

177

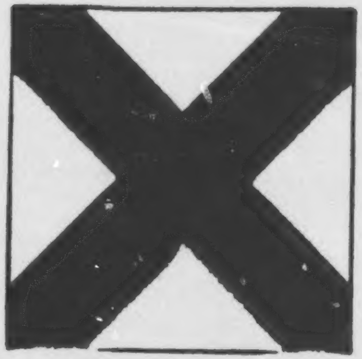


PATTERN

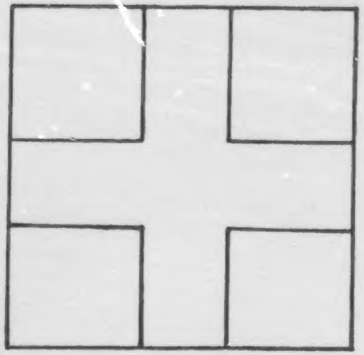
178



☐



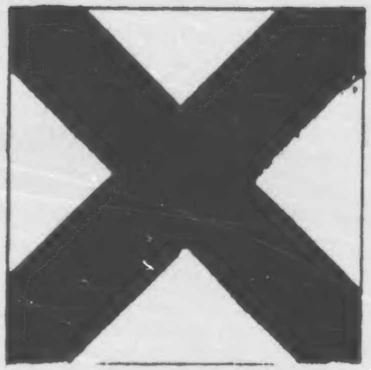
☐



☐

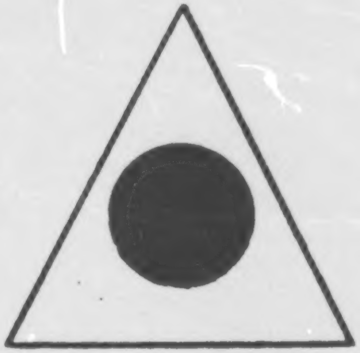
179

22

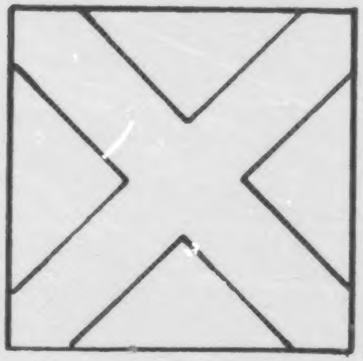


PATTERN

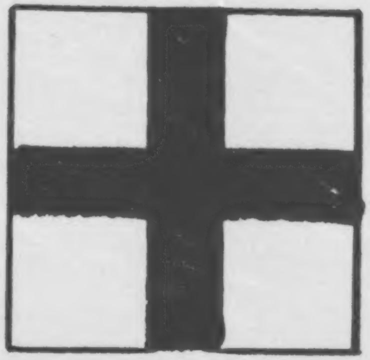
180



☐

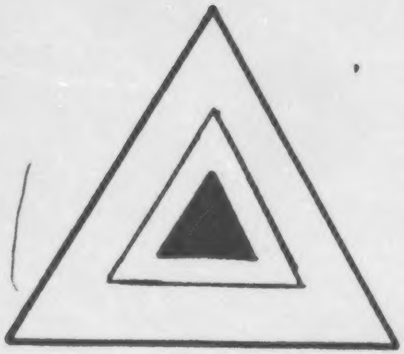


☐



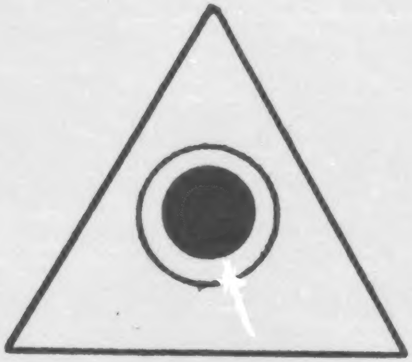
☐

181

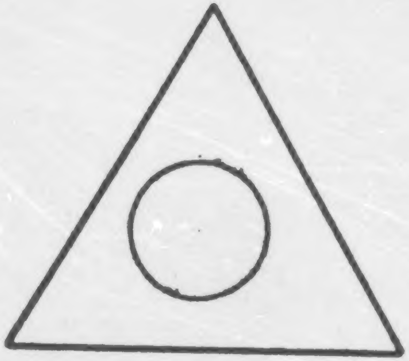


PATTERN

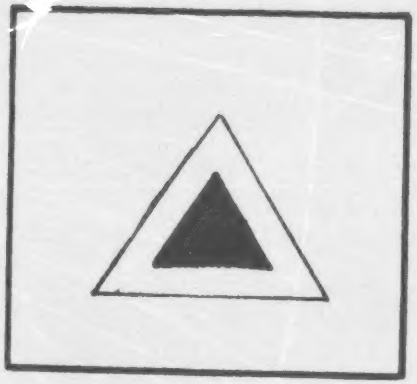
182



☐



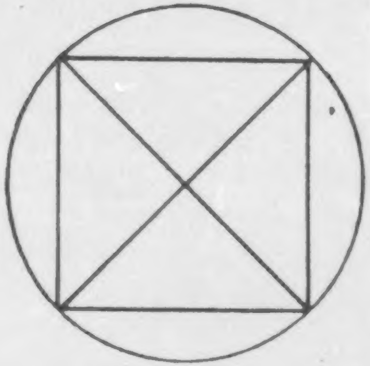
☐



☐

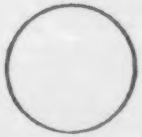
183

24

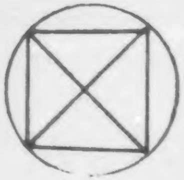


PATTERN

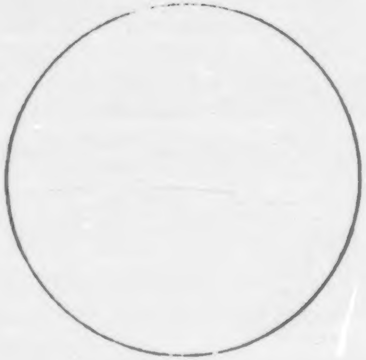
184



☐

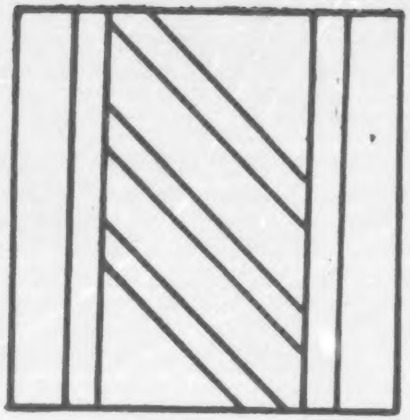


☐



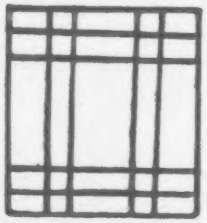
☐

185

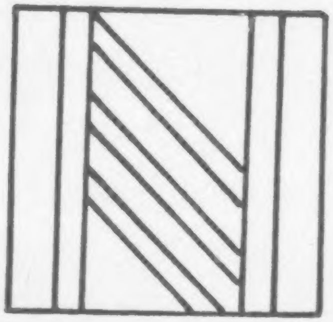


PATTERN

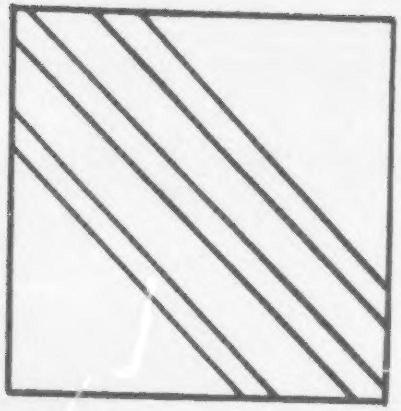
186



☐



☐



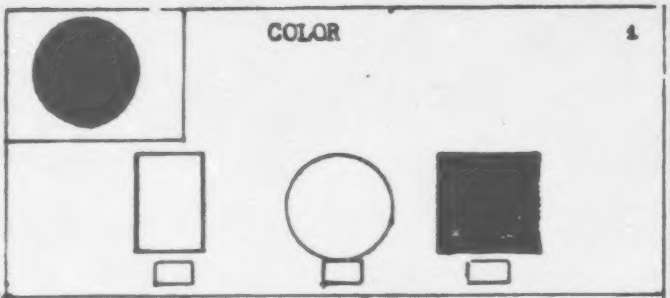
☐

187

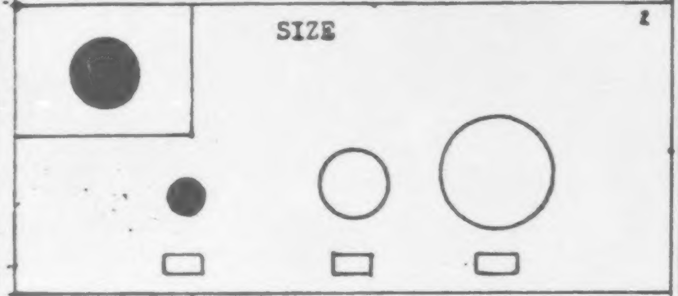
EXERCISE I: CRITERION MASTERY TEST

NAME (ID)

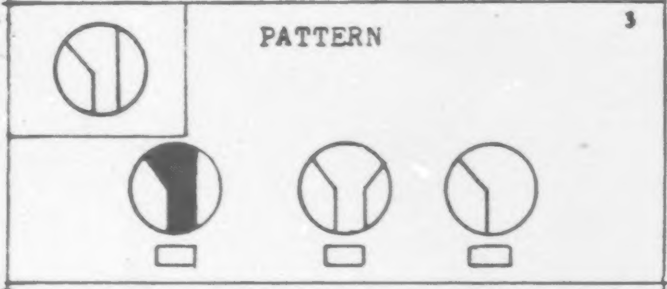
COLOR 1



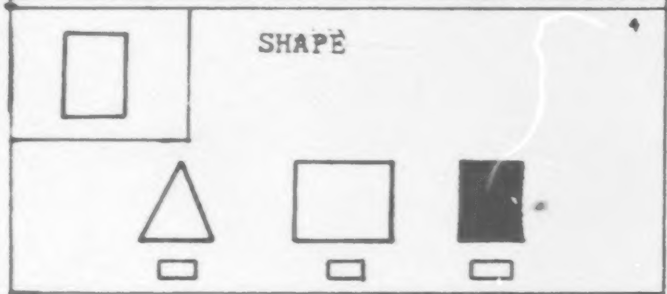
SIZE 2



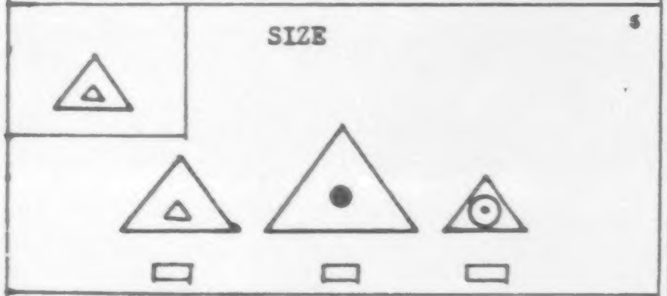
PATTERN 3



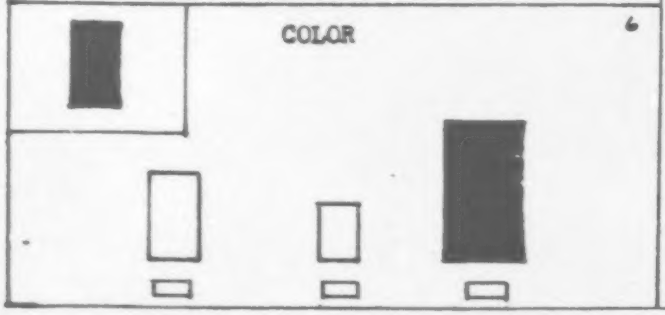
SHAPE 4



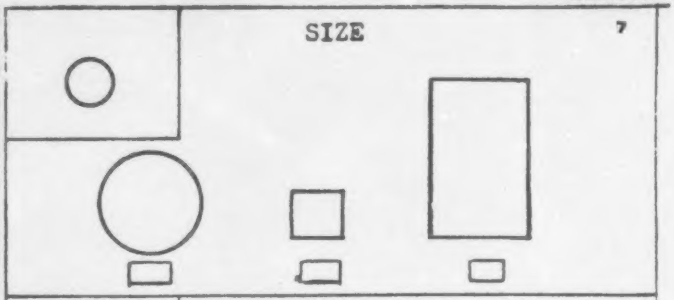
SIZE 5



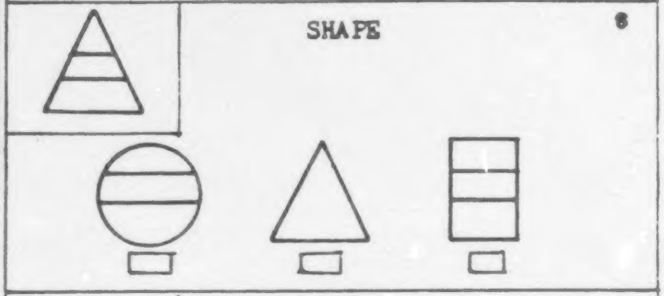
COLOR 6



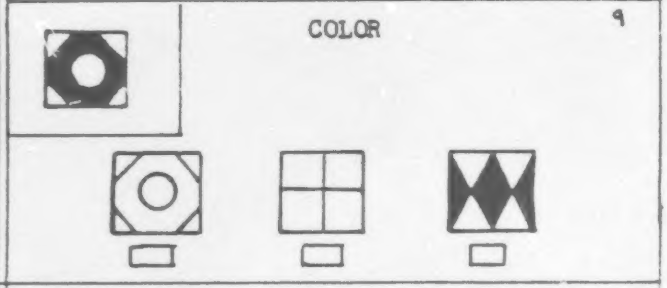
SIZE 7



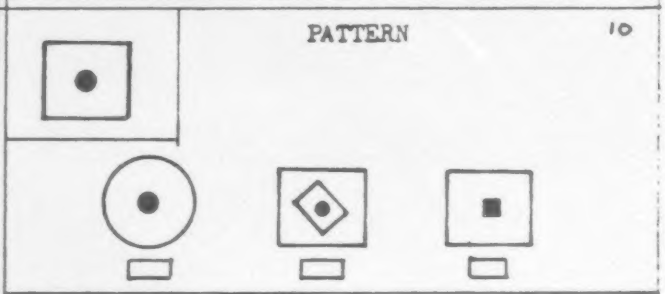
SHAPE 8



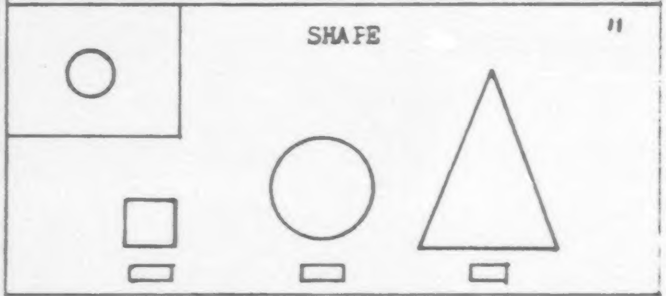
COLOR 9



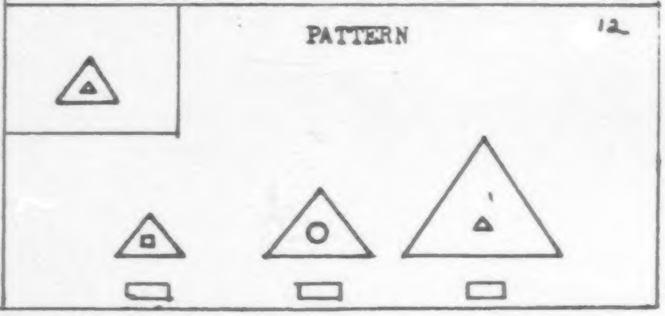
PATTERN 10



SHAPE 11



PATTERN 12

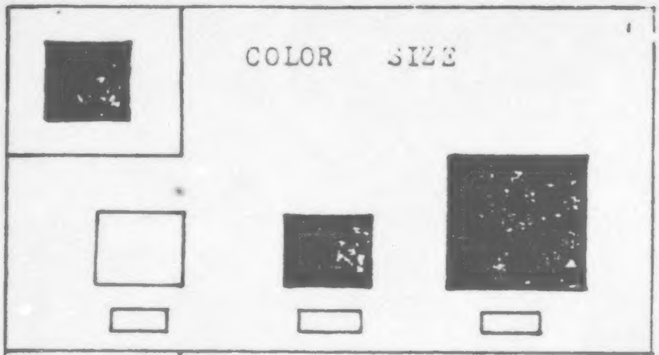


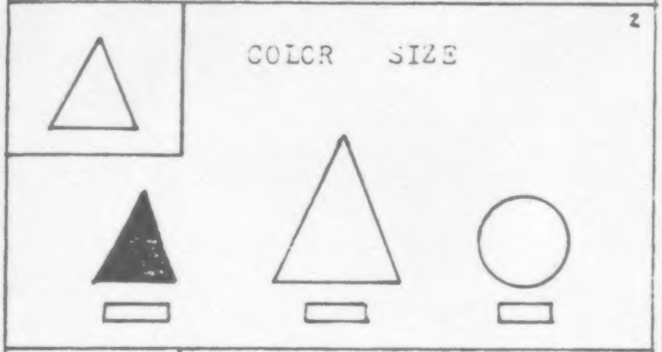
(F)

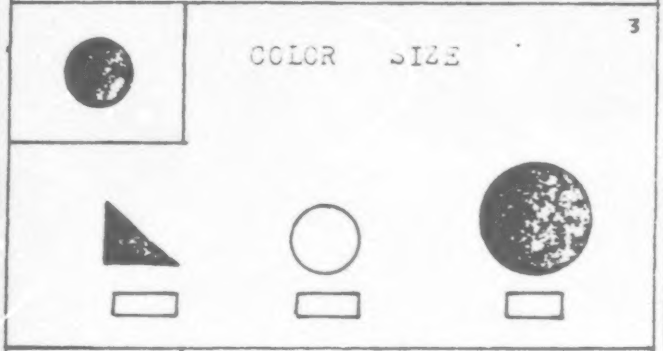
School

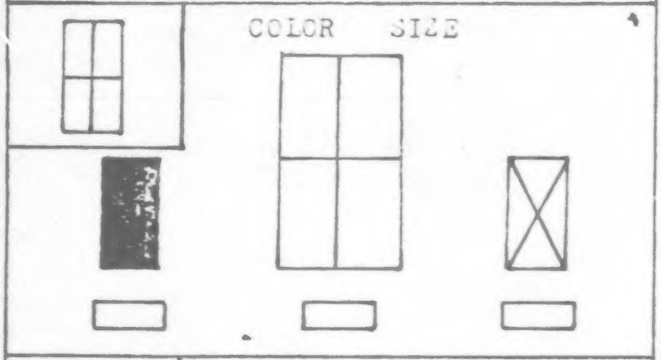
EXERCISE 2: TWO DIMENSIONS

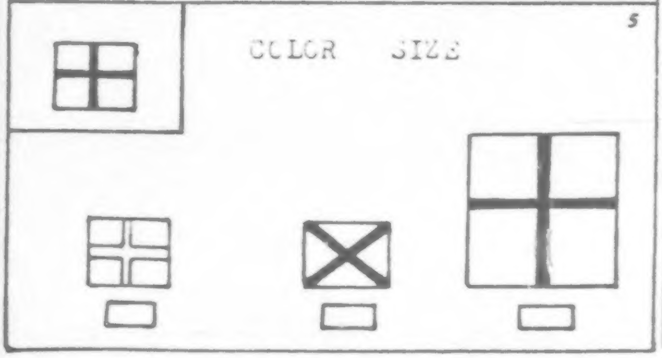
NAME (I.D.) ______________________

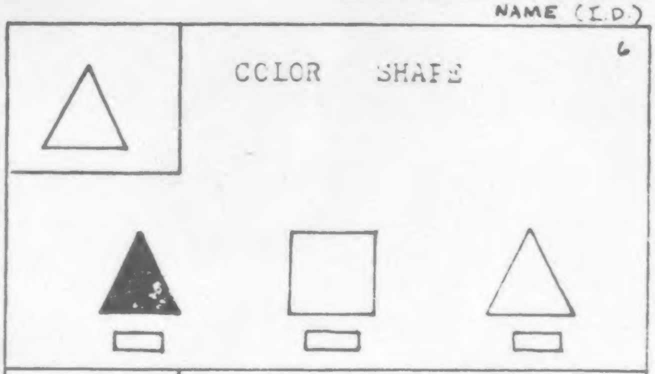

1 COLOR SIZE

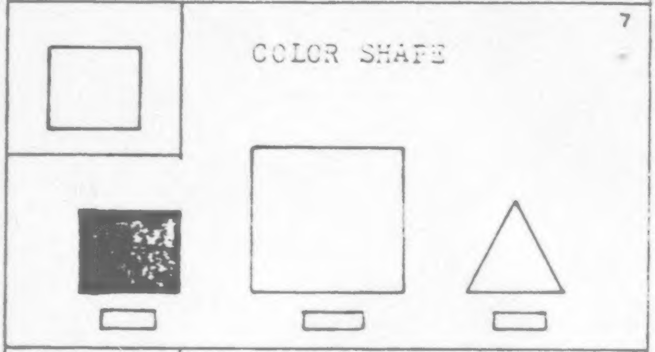

2 COLOR SIZE

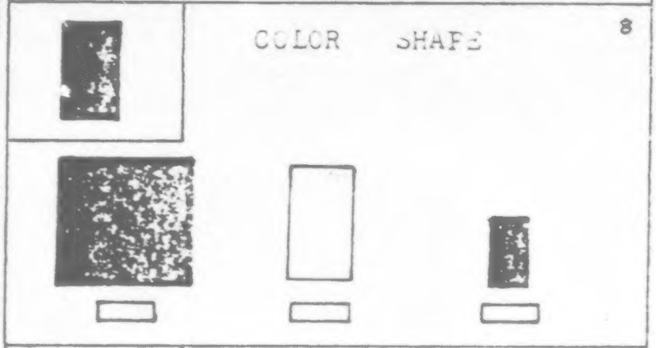

3 COLOR SIZE

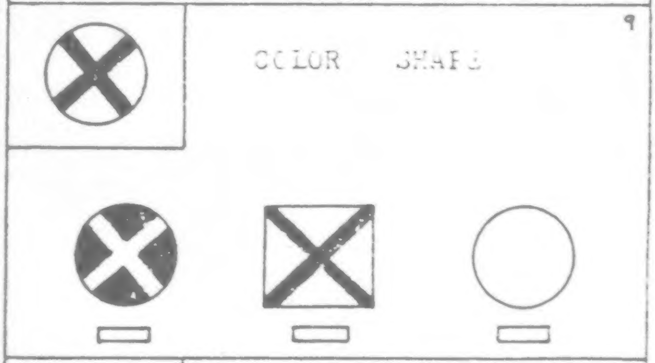

4 COLOR SIZE

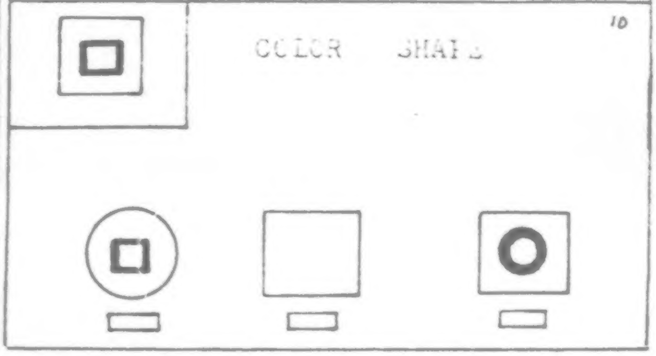

5 COLOR SIZE


6 COLOR SHAPE


7 COLOR SHAPE


8 COLOR SHAPE


9 COLOR SHAPE

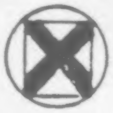

10 COLOR SHAPE

____
Grade

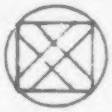
EXERCISE 2

Name (10) ____________________

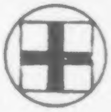
11



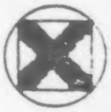
COLOR PATTERN



☐

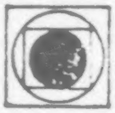


☐

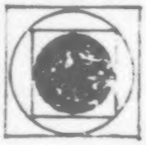


☐

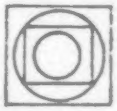
12



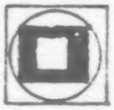
COLOR PATTERN



☐

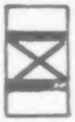


☐

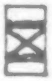


☐

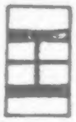
13



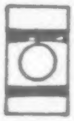
COLOR PATTERN



☐

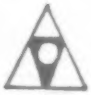


☐

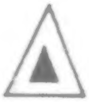


☐

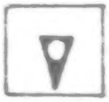
14



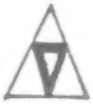
COLOR PATTERN



☐

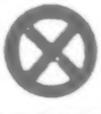


☐

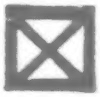


☐

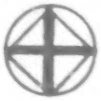
15



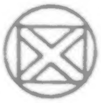
COLOR PATTERN



☐

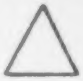


☐

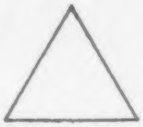


☐

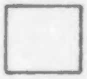
16



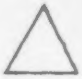
SHAPE SIZE



☐

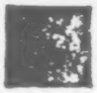


☐

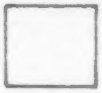


☐

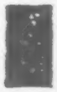
17



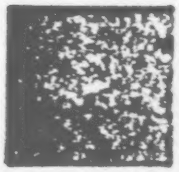
SHAPE SIZE



☐

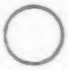


☐

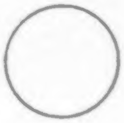


☐

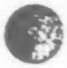
18



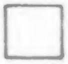
SHAPE SIZE



☐

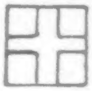


☐

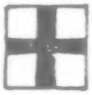


☐

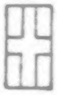
19



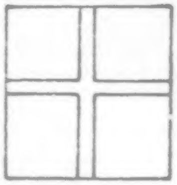
SHAPE SIZE



☐

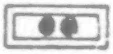


☐

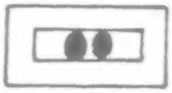


☐

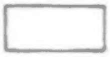
20



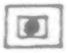
SHAPE SIZE



☐



☐

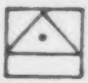


☐

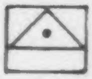
School ____________________

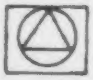
EXERCISE 2 ________________

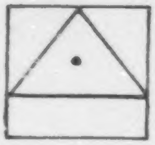
21



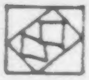
SIZE PATTERN


☐

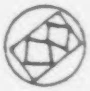

☐

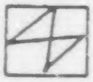

☐

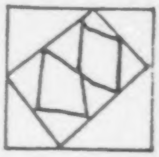
22



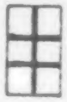
SIZE PATTERN


☐

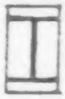

☐

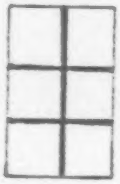

☐

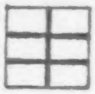
23



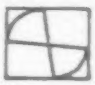
SIZE PATTERN


☐

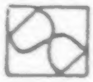

☐

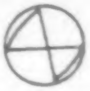

☐

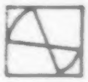
24



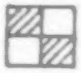
SHAPE PATTERN


☐

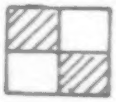

☐

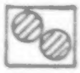

☐

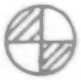
25



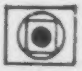
SHAPE PATTERN


☐

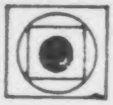

☐

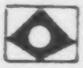

☐

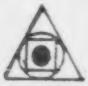
26



SHAPE PATTERN


☐


☐


☐

☐ ☐ ☐

☐ ☐ ☐

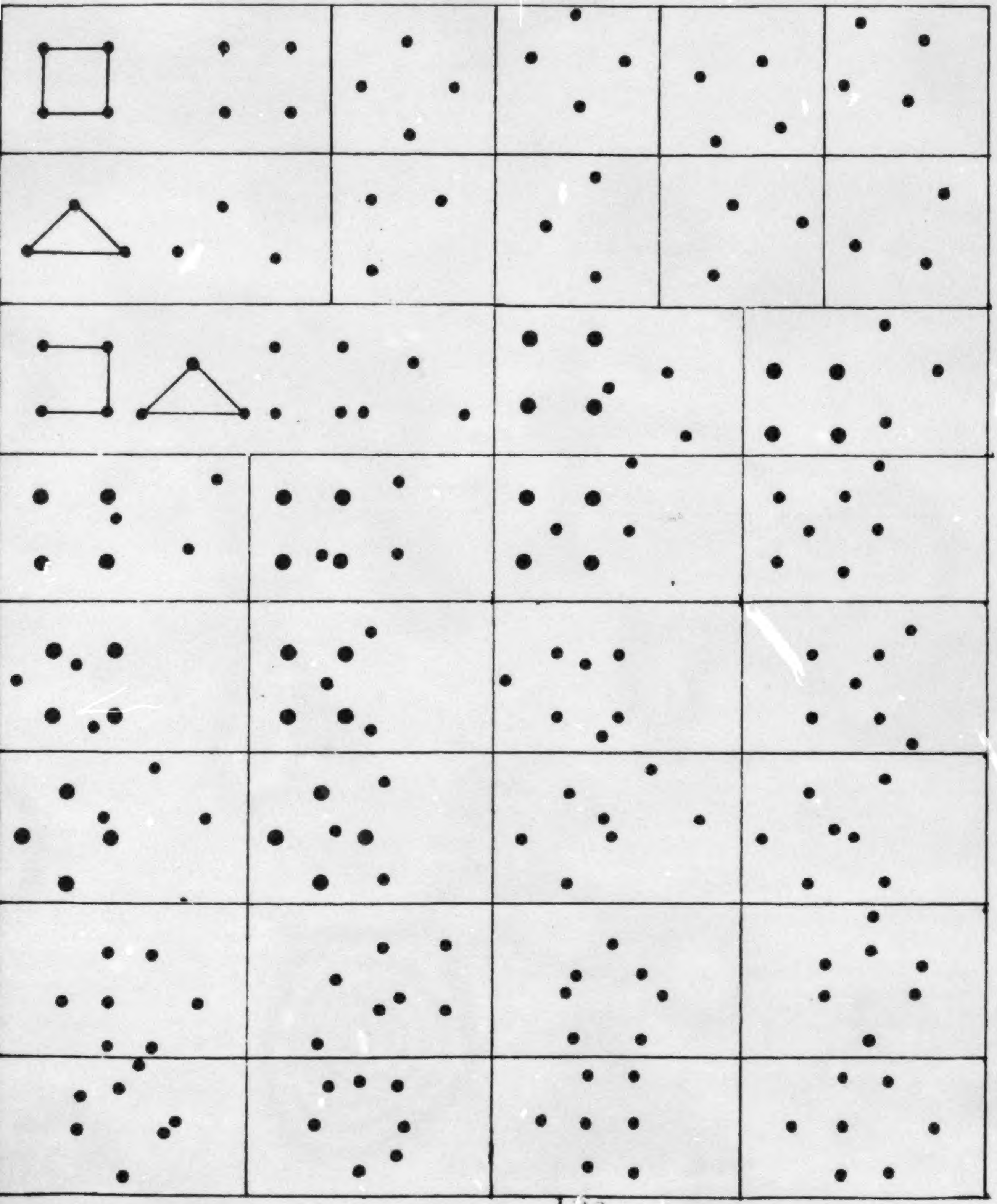
☐ ☐ ☐

☐ ☐ ☐

NIE STUDY : Dots Training Sheet, EXERCISE 9

Name: ______________________________ I.D. ______________

School: ______________________ Grade: ________ Date: ______________________

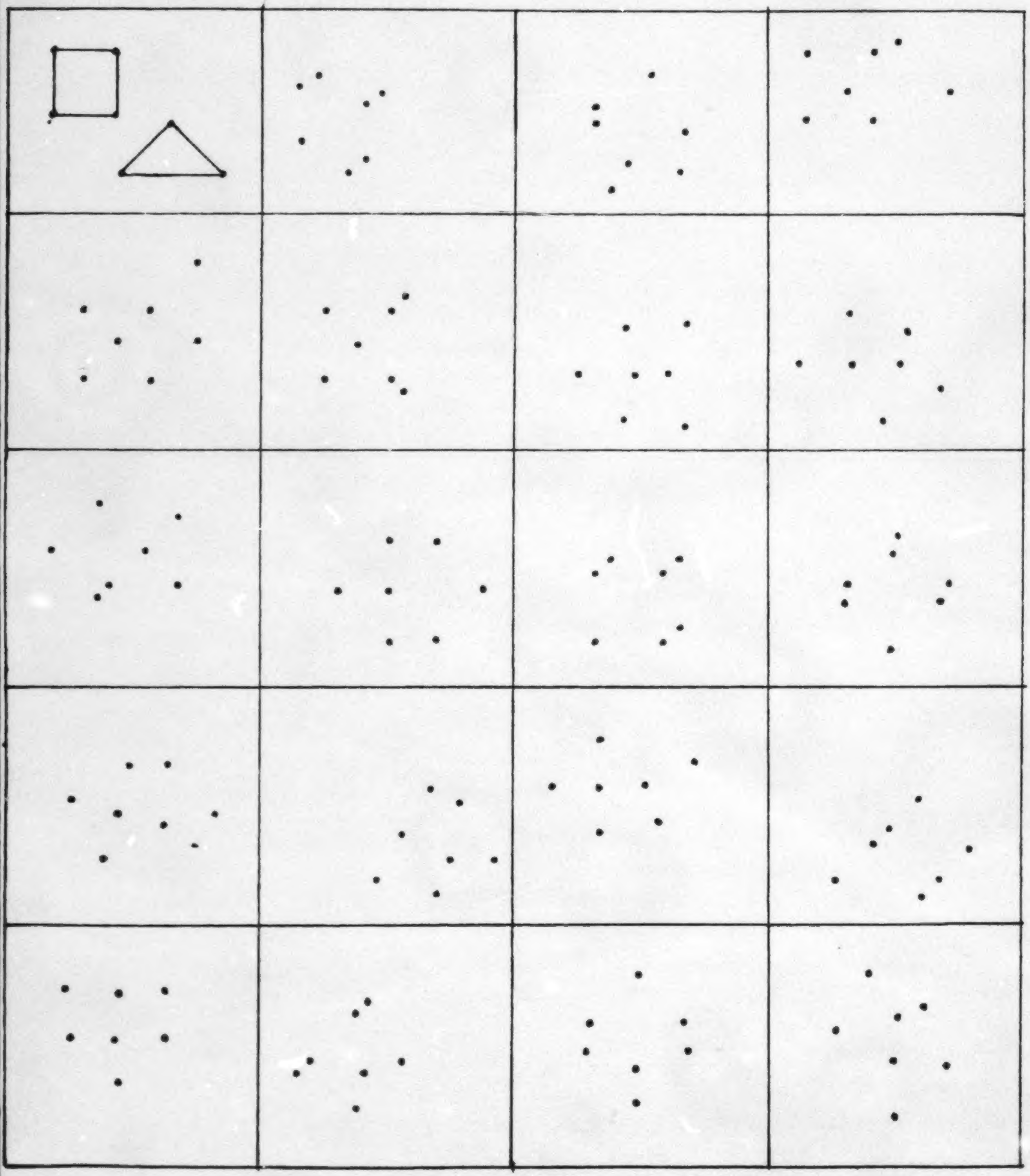


Exercise 9: Dot's Criterion mastery Test

Directions: Connect the dots to form one triangle and one square. Each dot belongs to only one shape.

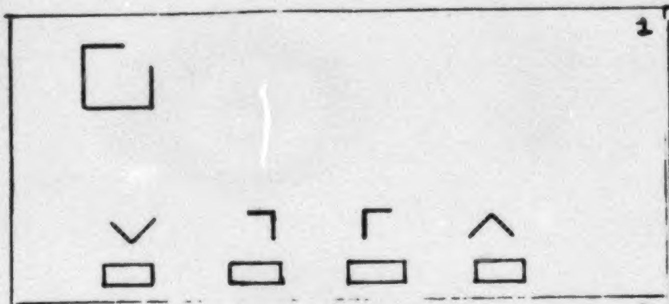
Name ______________________

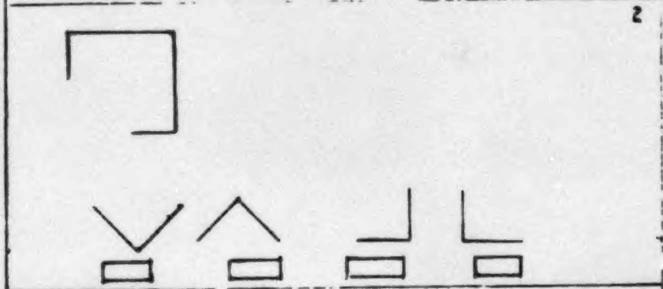
School ______________ GRADE ________ Date ________

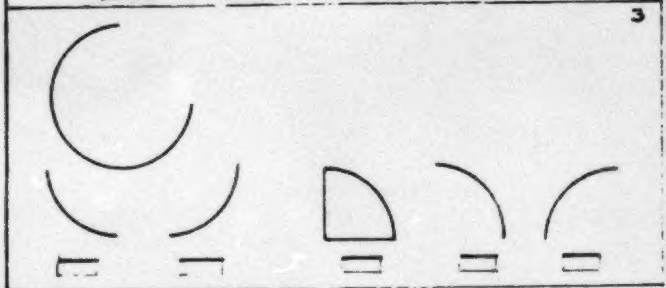


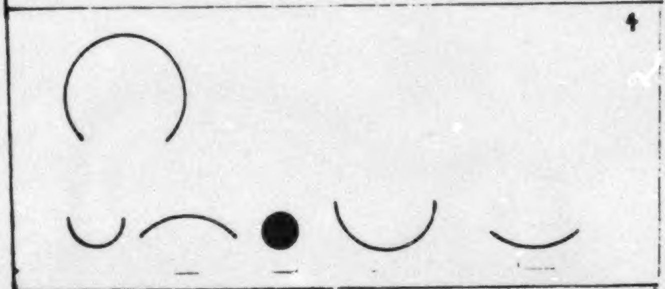
Exercise 5: Figure Completion

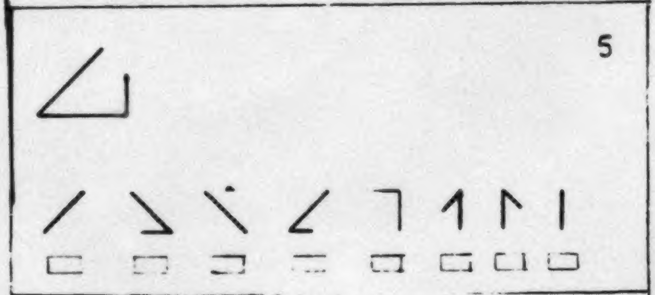
Name (I.D.)

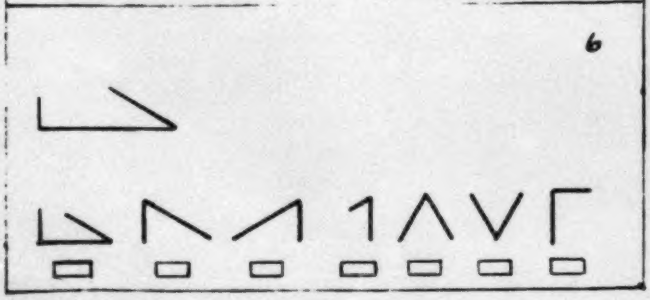


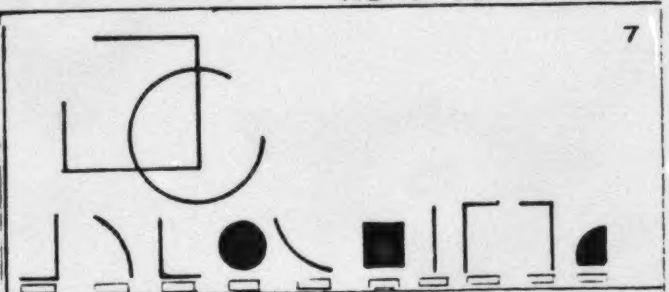


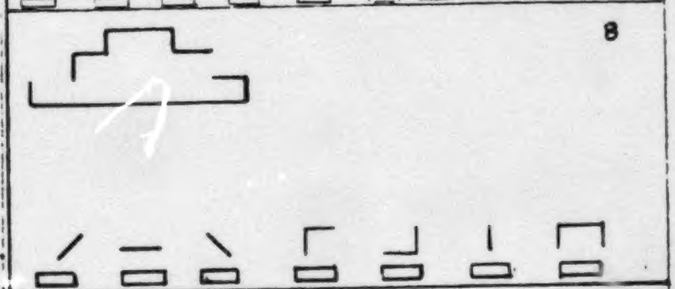


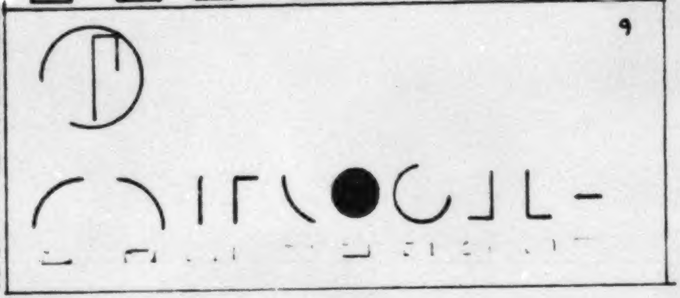


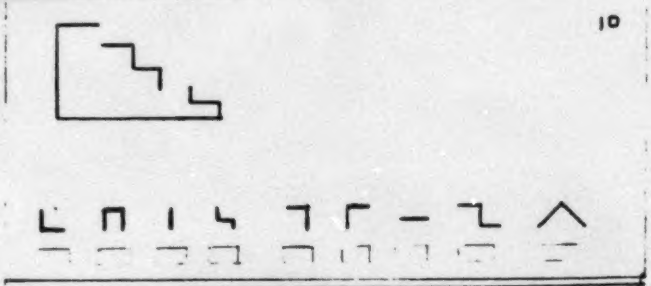


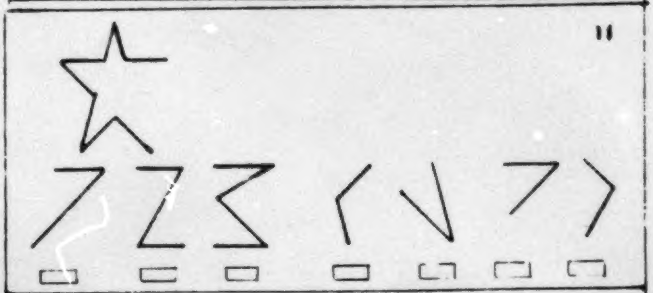


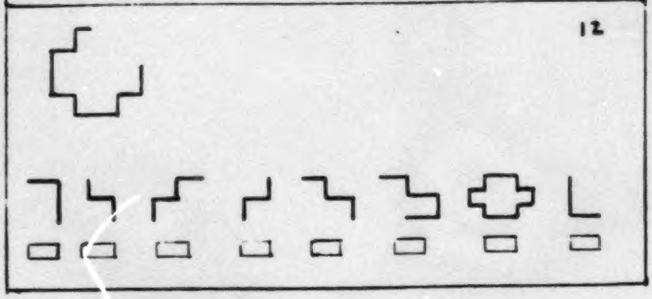










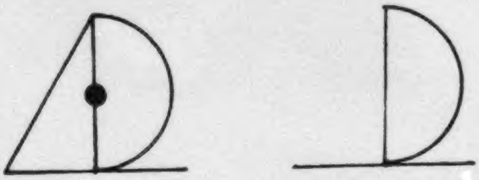


School

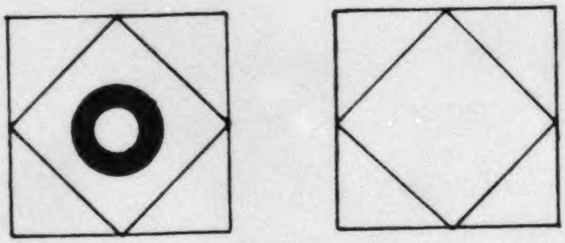
Exercise 6: Pattern Completion

Name ____________________

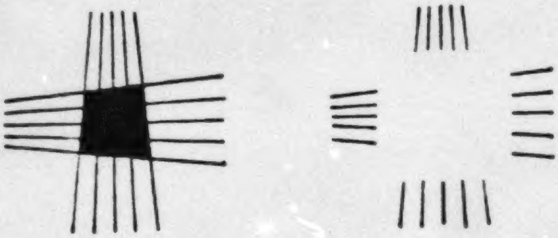
School ____________________



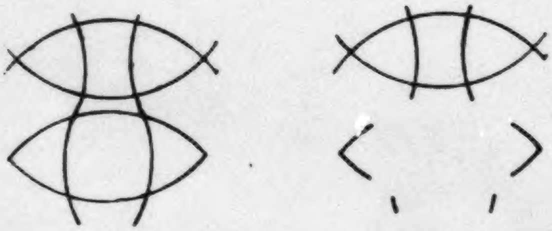
1



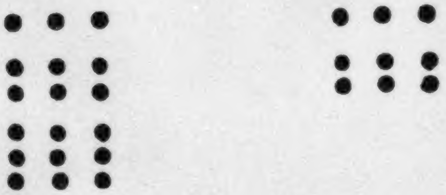
2



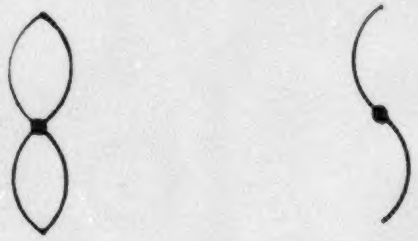
3



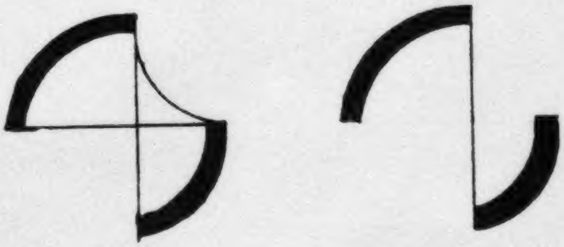
4



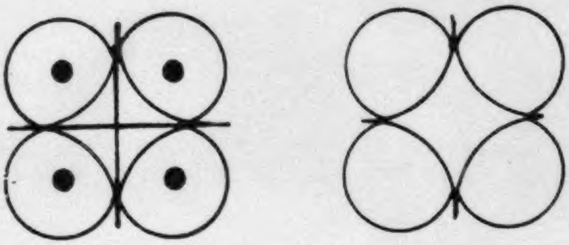
5



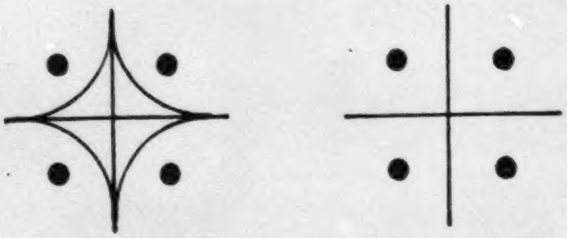
6



7



8

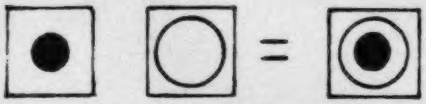


Exercise 7: Combining Patterns

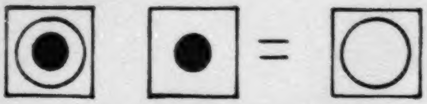
Name

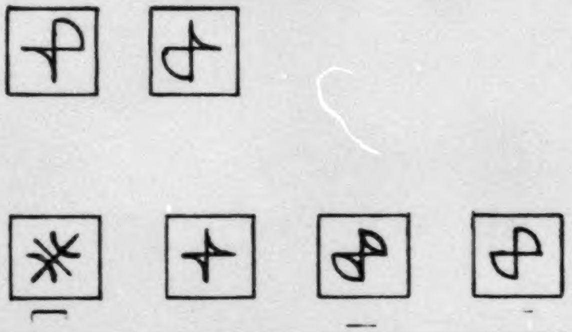
School

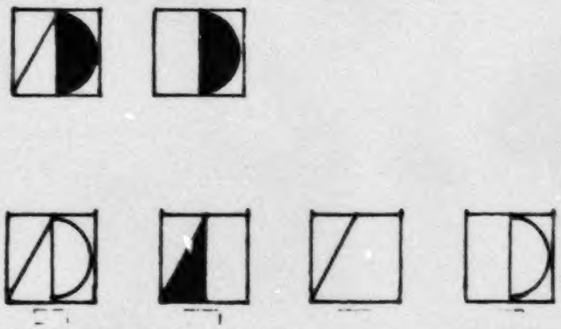
Adding Patterns:

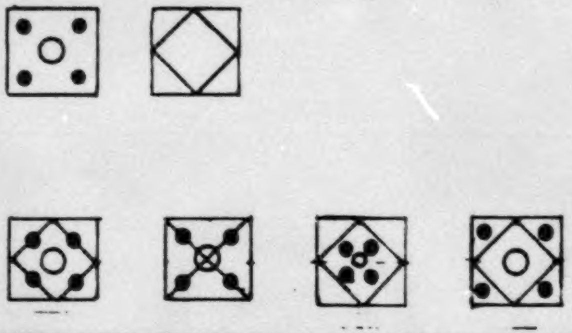


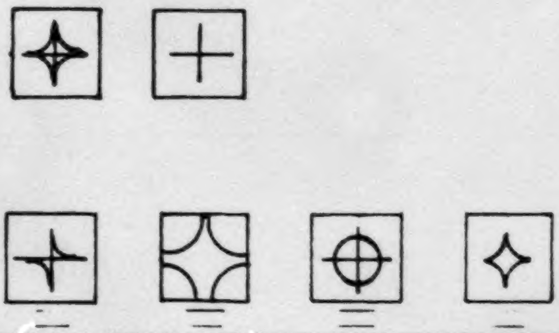
Subtracting Patterns:

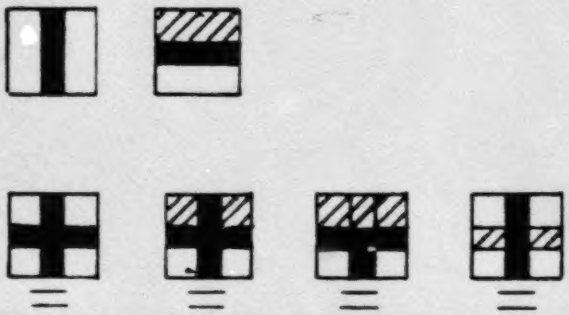


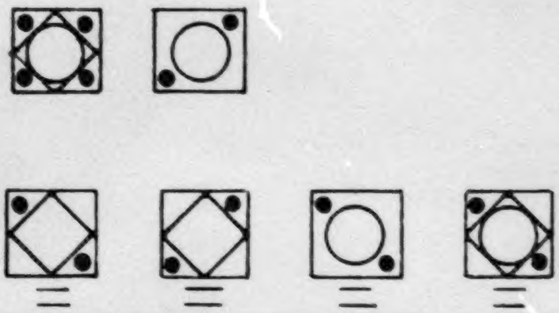


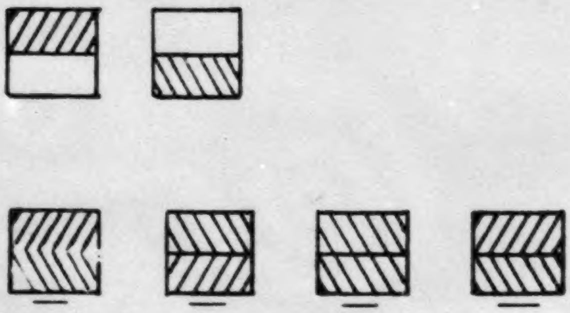


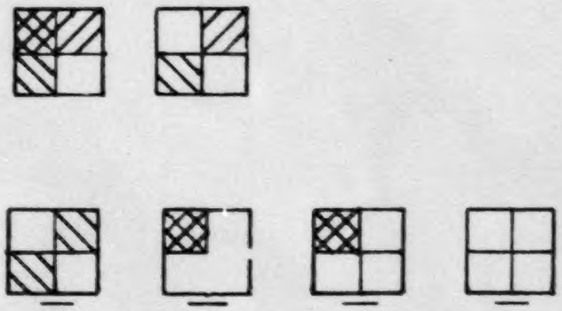








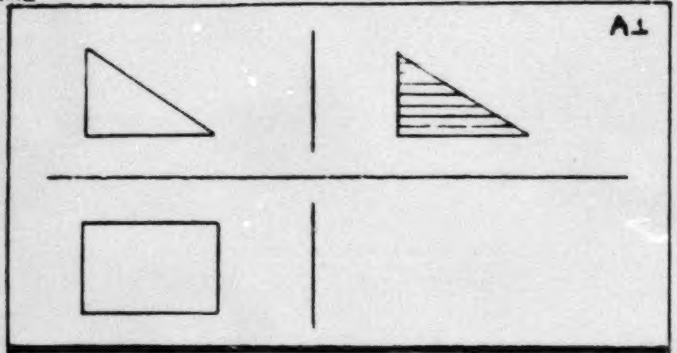




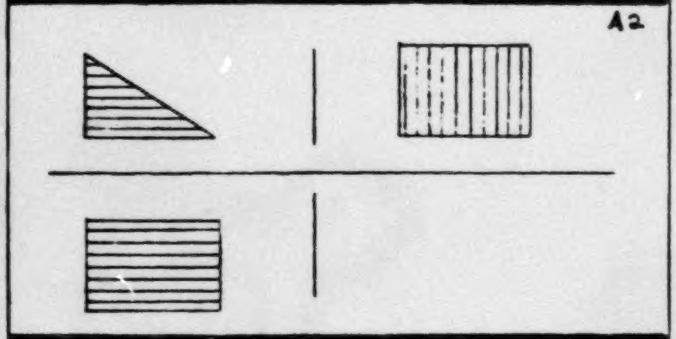
EXERCISE: 8

AME ____________________ SCHOOL ____________________

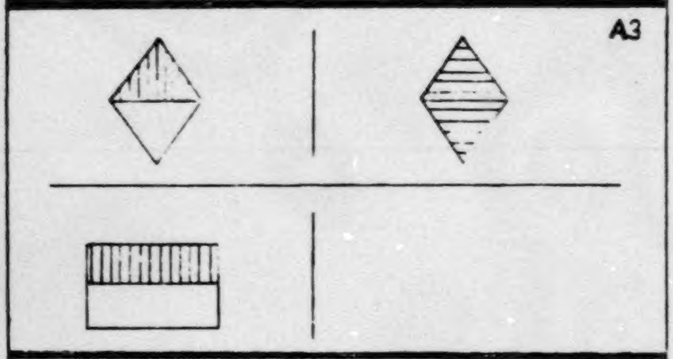
A1



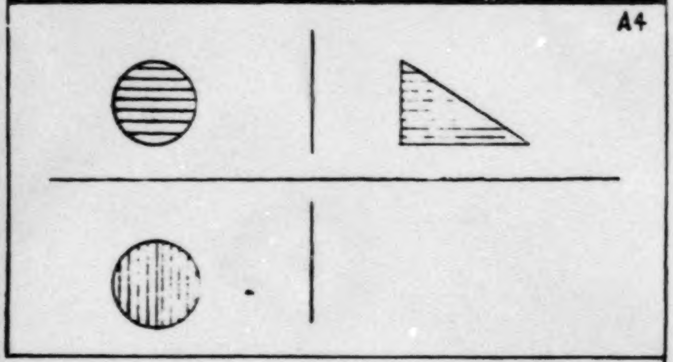
A2



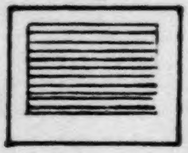
A3



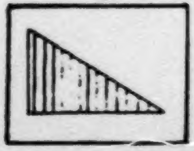
A4



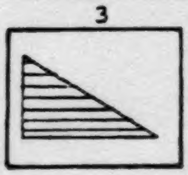
1



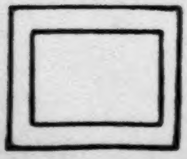
2



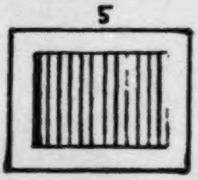
3



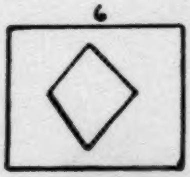
4



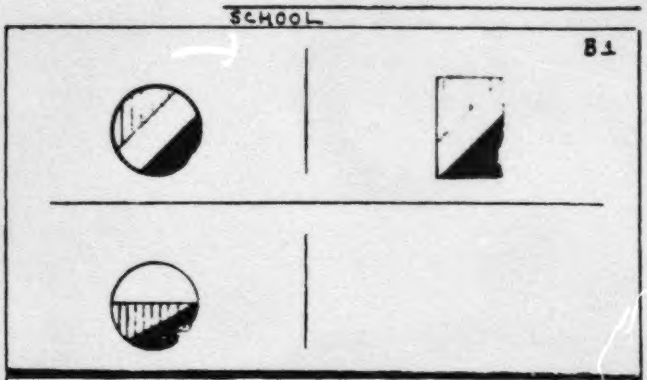
5



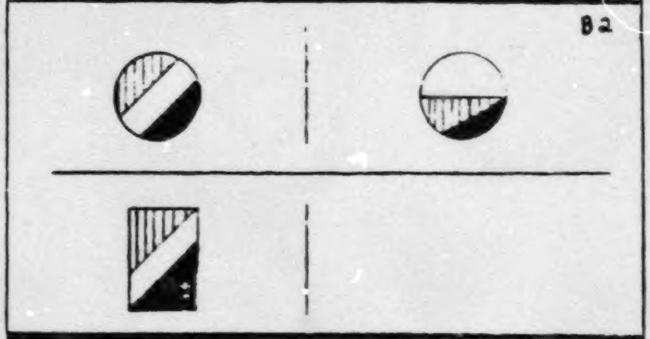
6



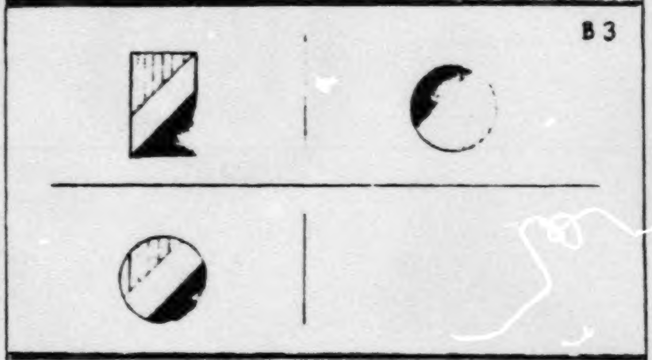
B1



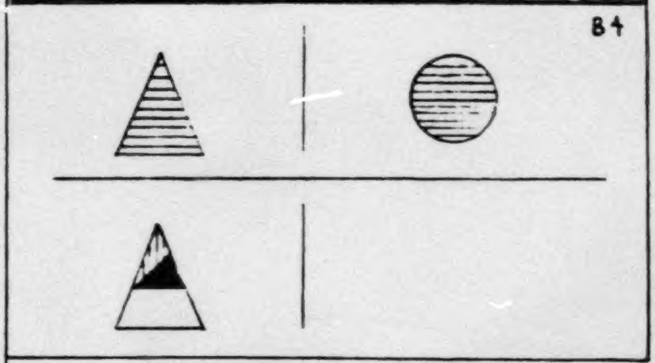
B2



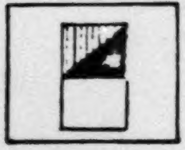
B3



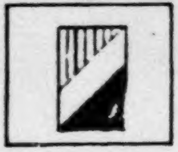
B4



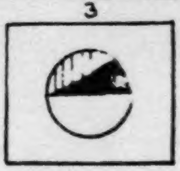
1



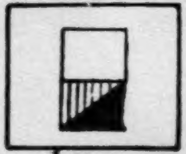
2



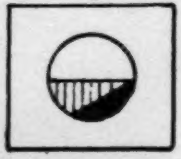
3



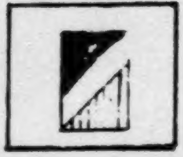
4



5



6

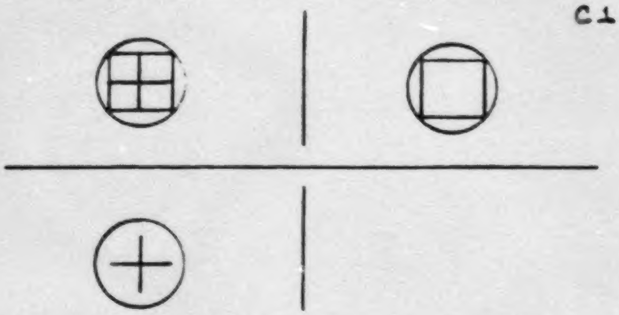


EXERCISE: 8

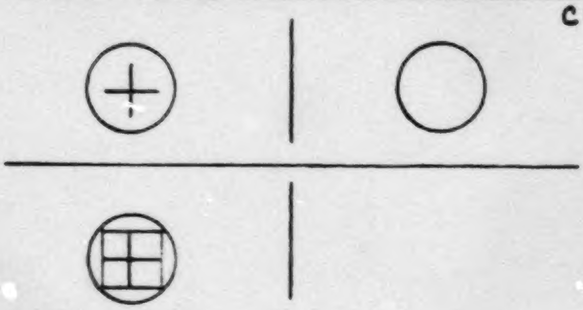
ME

SCHOOL

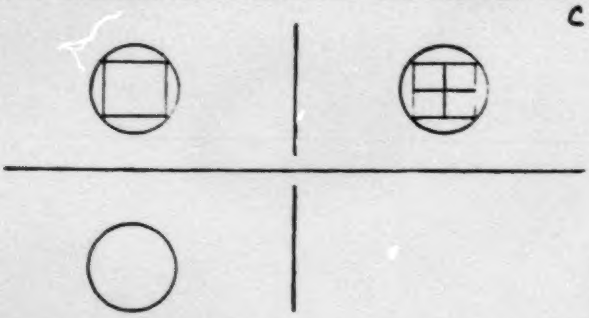
C1



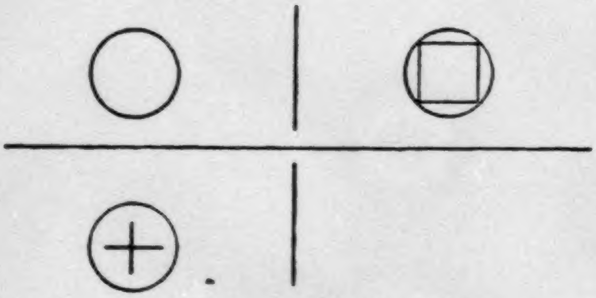
C2



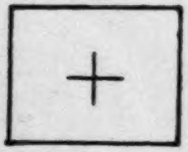
C3



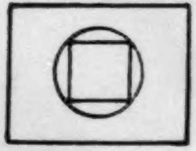
C4



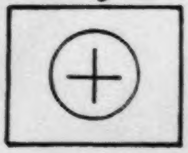
1



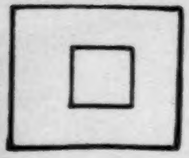
2



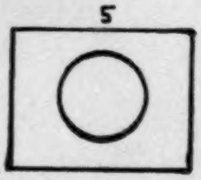
3



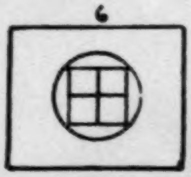
4



5



6



1

2

3

4

1

2

3

4

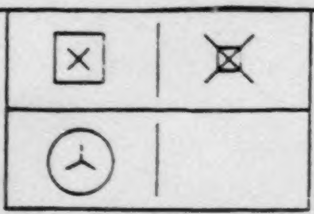
5

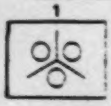
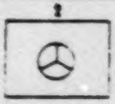
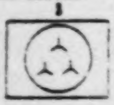
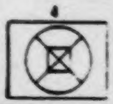
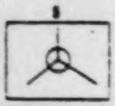
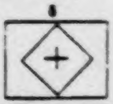
6

Exercise 9: Analogies Criterion Test

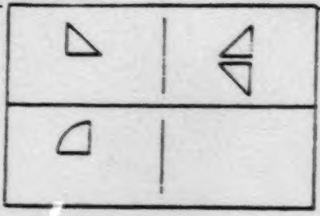
Name

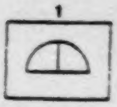
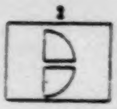
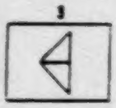
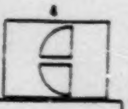
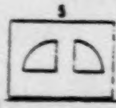
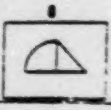
1



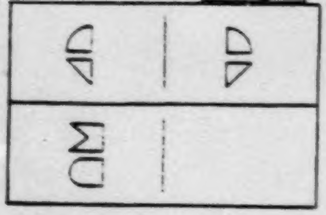
1

2

3

4

5

6


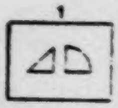
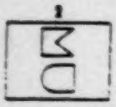
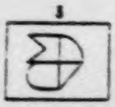
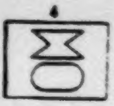
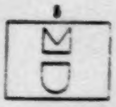
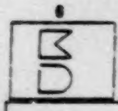
2



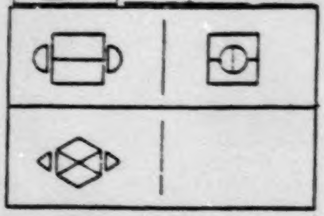
1

2

3

4

5

6


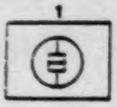
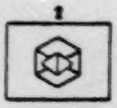
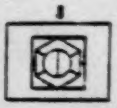
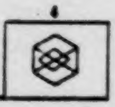
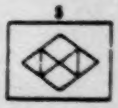
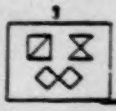
3



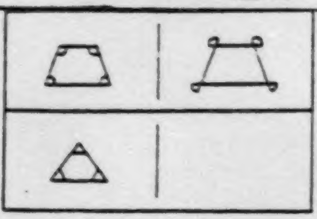
1

2

3

4

5

6


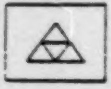
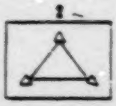
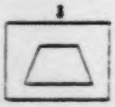
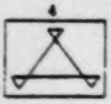
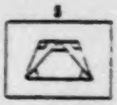
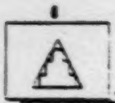
4



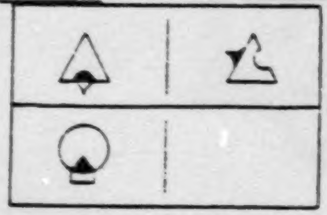
1

2

3

4

5

6


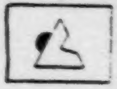
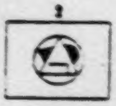
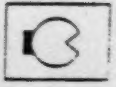
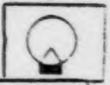
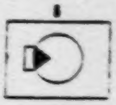
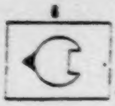
5



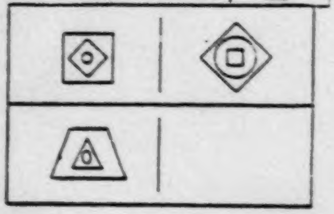
1

2

3

4

5

6


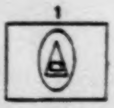
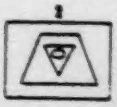
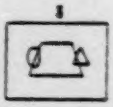
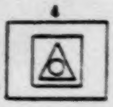
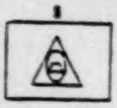
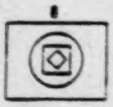
6



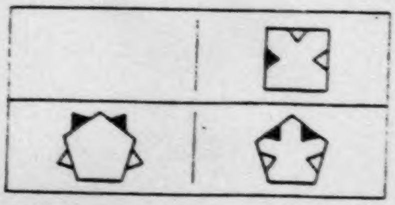
1

2

3

4

5

6


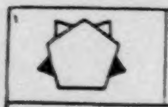
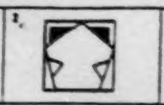
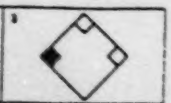
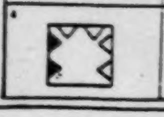
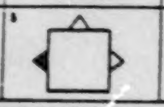
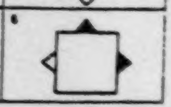
7



1

2

3

4

5

6


8



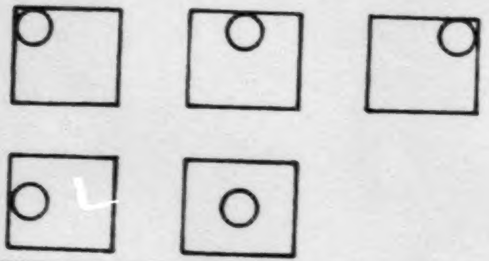
1

2

3

4

5

6


School

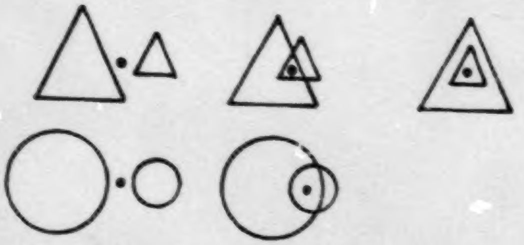
EXERCISE 10: 2×3 ANALOGIES

NAME

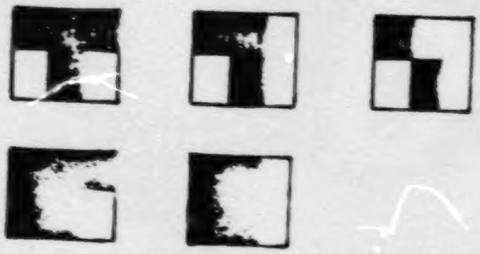
1



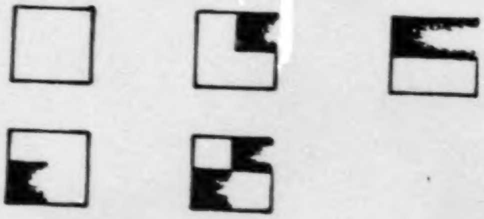
2



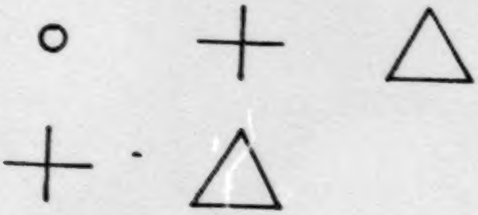
3



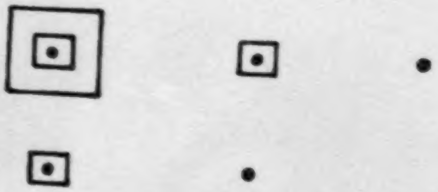
4



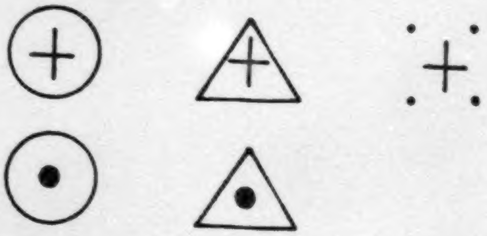
5



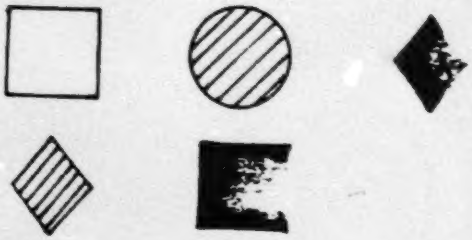
6



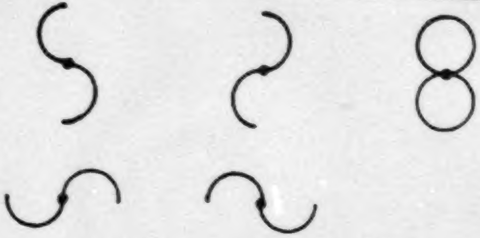
7



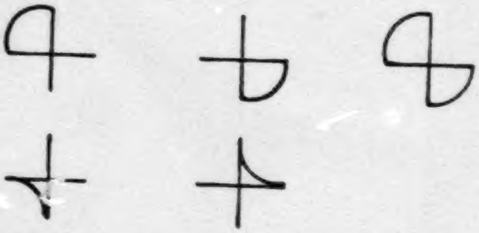
8



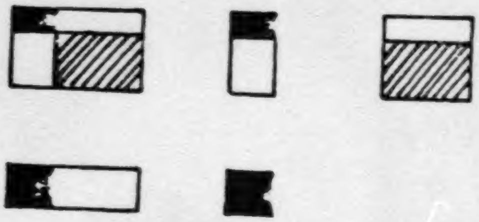
9



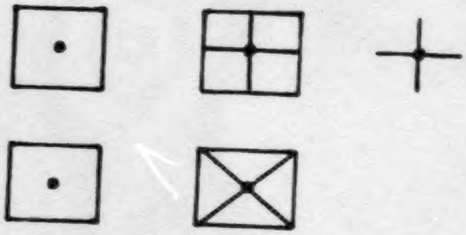
10



11



12



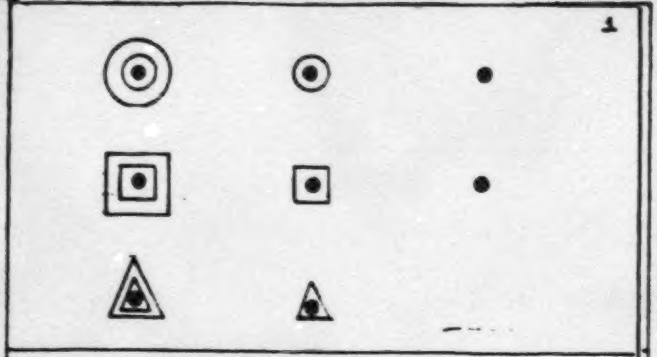
School

AME

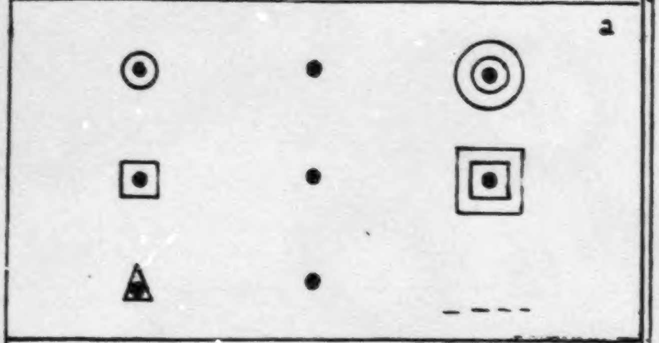
EXERCISE

SCHOOL

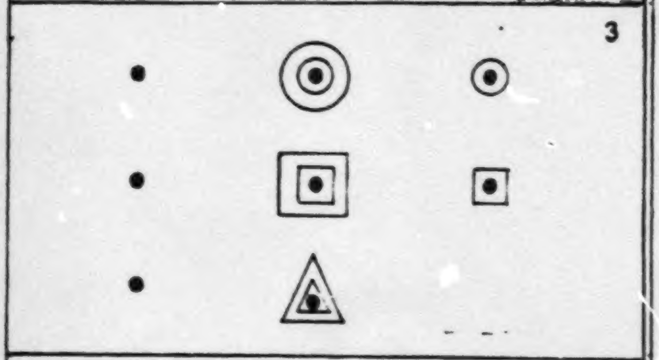
1



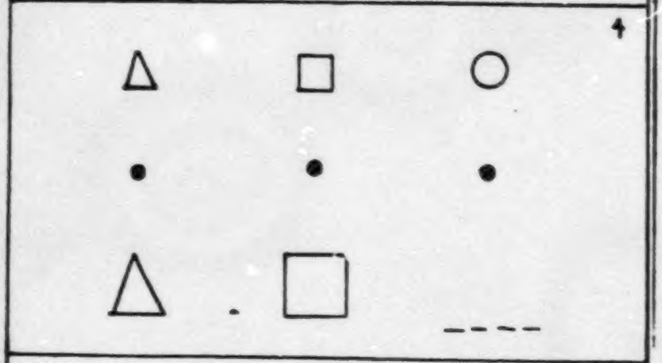
2



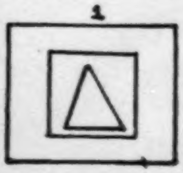
3



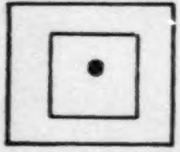
4



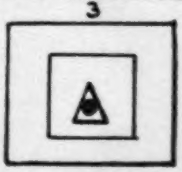
1



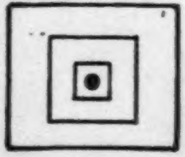
2



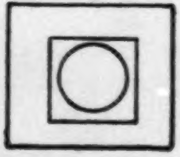
3



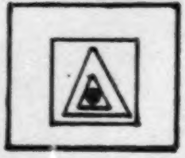
4



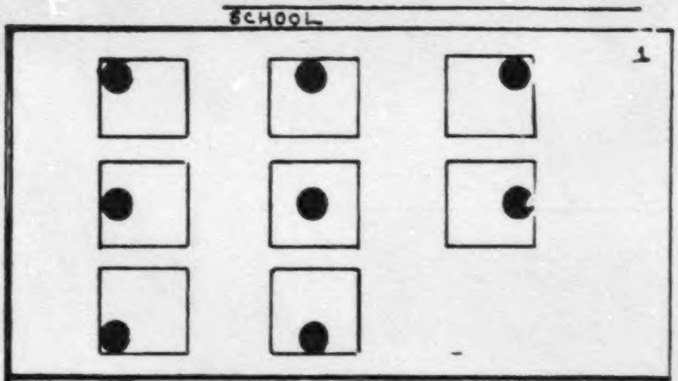
5



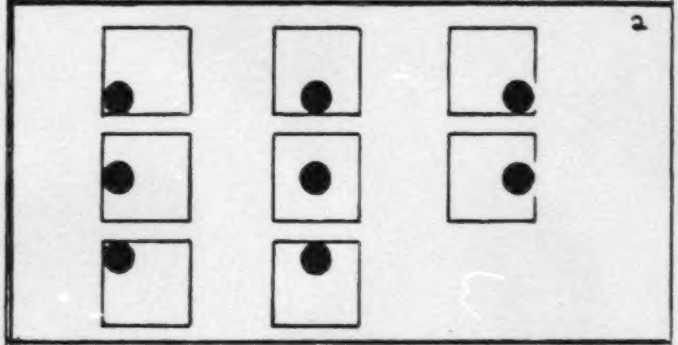
6



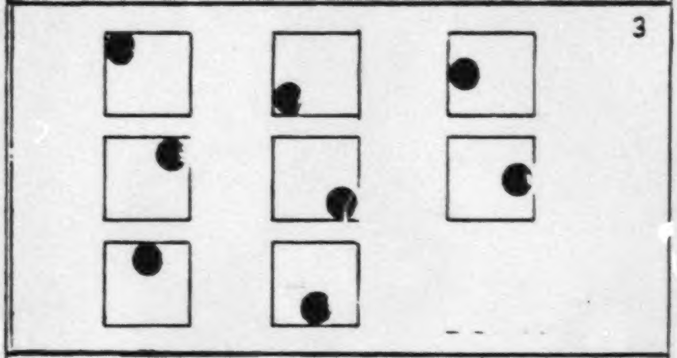
1



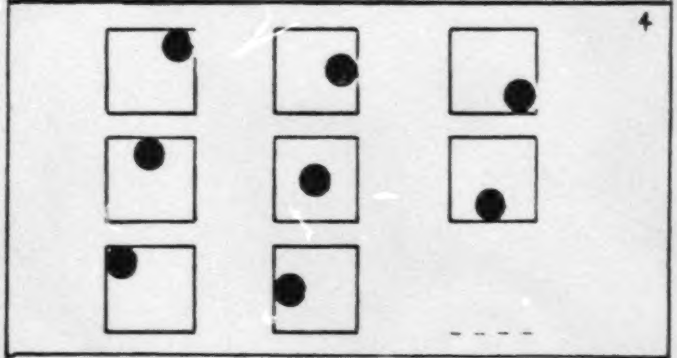
2



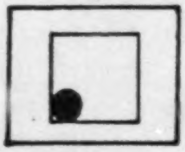
3



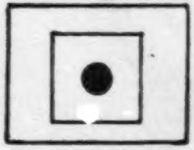
4



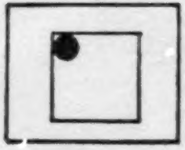
1



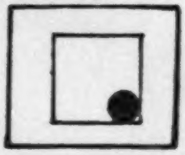
2



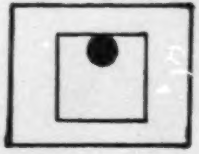
3



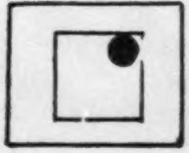
4



5



6

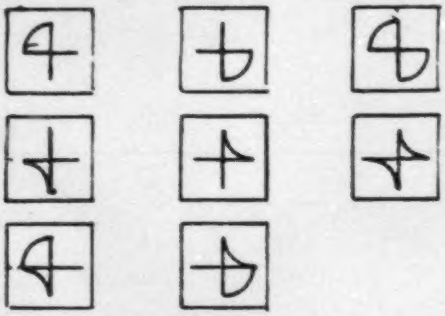


EXERCISE: 11

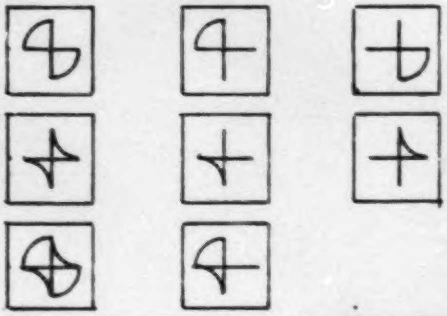
ME

SCHOOL

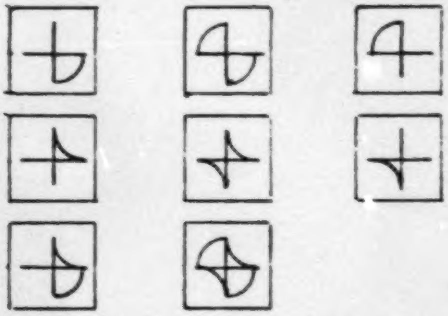
1



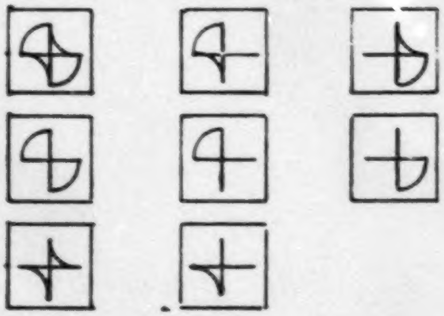
2



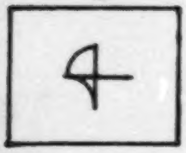
3



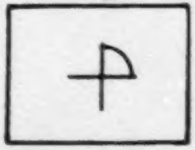
4



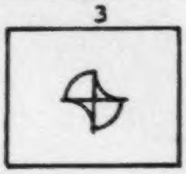
1



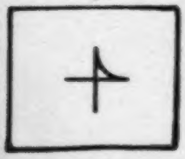
2



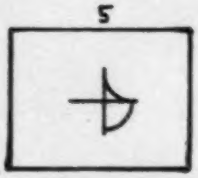
3



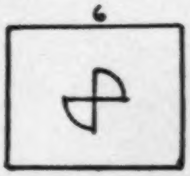
4



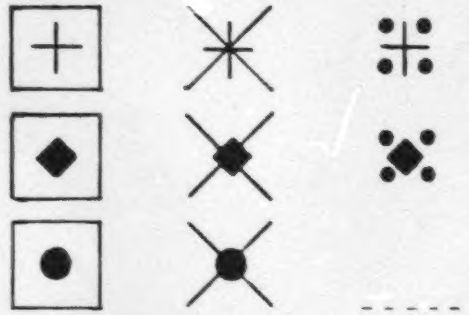
5



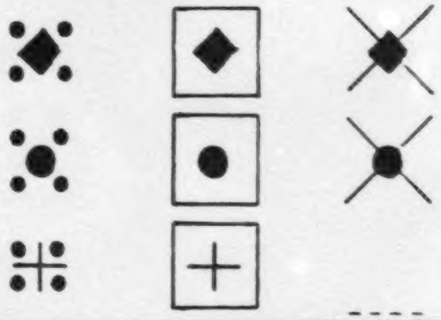
6



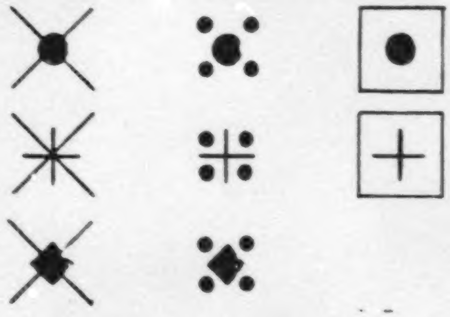
1



2

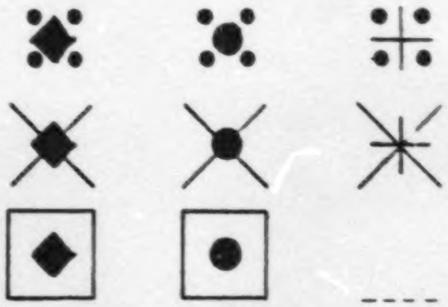


3

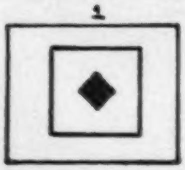


--

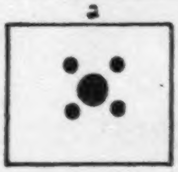
4



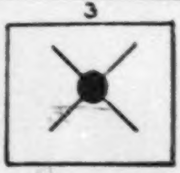
1



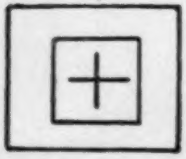
2



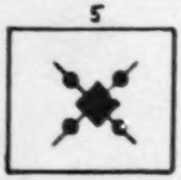
3



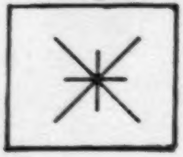
4



5



6



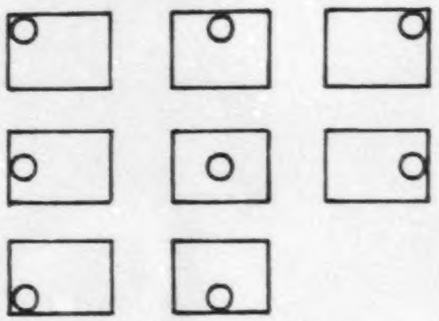
EXERCISE : 12

3X3 CRITERION MASTERY

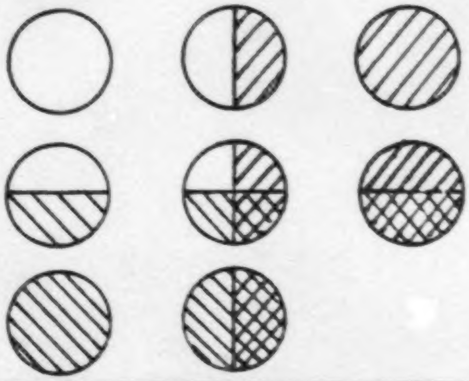
SCHOOL

NAME

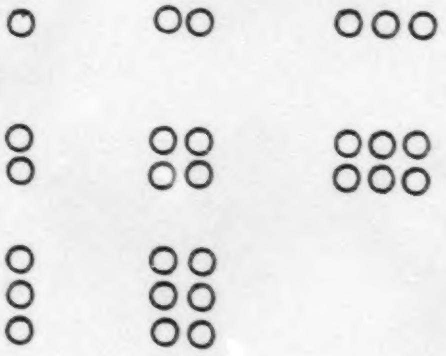
1



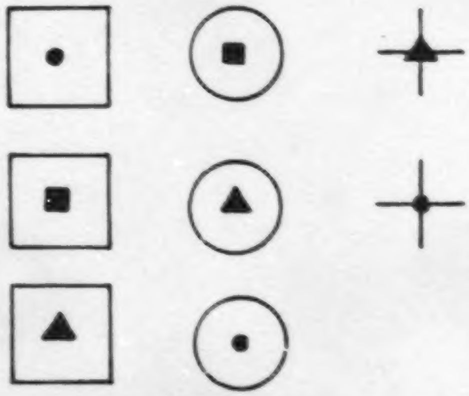
2



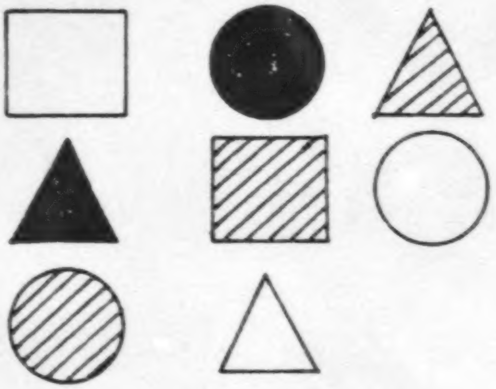
3



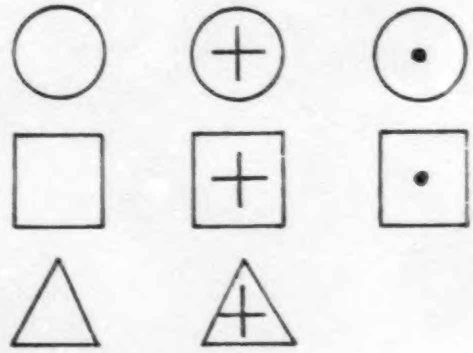
4



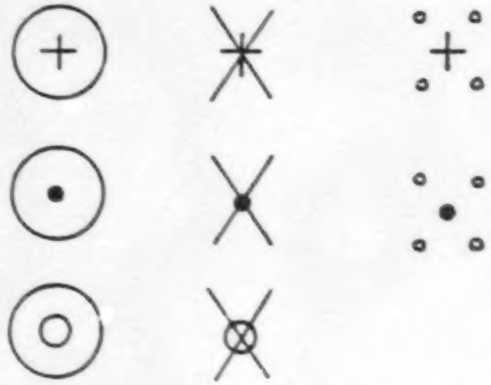
5



6



7



8